PCT/EP200 4 / 0 5 2 0 0 0

Europäisches Patentamt | **European Patent Office** | **Office européen des brevets**

02. 09. 2004

REC'D 27 OCT 2004
WIPO PCT

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03102699.0

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1 (a) OR (b)

Der Präsident des Europäischen Patentamts; Im Auftrag

For the President of the European Patent Off

Le Président de l'Office européen des brevet p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 03102699.0
Demande no:

Anmeldetag:
Date of filing: 04.09.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Applied Research Systems ARS Holding N.V.
Pietermaai 15
Curacao
ANTILLES NEERLANDAISES

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

NOVEL UBP8rp POLYPEPTIDES AND THEIR USE IN THE TREATMENT OF PSORIASIS

In Anspruch genommene Prioriät(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C07K14/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI

NOVEL UBP8rp POLYPEPTIDES AND THEIR USE IN THE TREATMENT OF PSORIASIS

FIELD OF THE INVENTION

This invention relates to a novel gene encoding a protein of the ubiquitin-proteasome pathway, UBP8rp. The invention also relates the use of UBP8rp polypeptides for screening for modulators, and to the use of said modulators for treating chronic inflammatory diseases such as, e.g., psoriasis, psoriatic arthritis, rheumatoid arthritis, asthma, inflammatory bowel disease and multiple sclerosis. The invention further relates to the use of biallelic markers located in the UBP8rp gene for diagnosing said chronic inflammatory diseases.

BACKGROUND

1. Psoriasis

Psoriasis is a chronic, recurring disease recognizable by silvery scaling bumps and various - sized plaques (raised patches). An abnormally high rate of growth and turnover of skin cells causes the scaling. The reason for the rapid cell growth is unknown, but immune mechanisms are thought to play a role. The condition often runs in families. Psoriasis is common, affecti ng 2 to 4 % of whites, blacks are less likely to get the disease. Psoriasis begins most often in people aged 10 to 40, although people in all age groups are susceptible.

1.1. Symptoms

Psoriasis usually starts as one or more small psoriatic plaques that be come excessively flaky. Small bumps may develop around the area. Although the first plaques may clear up by themselves, others may soon follow. Some plaques may remain thumbnail -sized, but others may grow to cover large areas of the body, sometimes in stri king ring-shaped or spiral patterns.

Psoriasis typically involves the scalp, elbows, knees, back, and buttocks. The flaking may be mistaken for severe dandruff, but the patchy nature of psoriasis, with flaking areas interspersed among completely normal ones, distinguishes the disease from dandruff. Psoriasis can also break out around and under the nails, making them thick and deformed. The eyebrows, armpits, navel, and groin may also be affected. Usually, psoriasis produces only flaking. Even itching is uncommon. When flaking areas heal, the skin takes on a completely normal appearance, and hair growth is unchanged. Most people with limited psoriasis suffer few problems beyond the flaking, although the skin's appearance may be embarrassing.

Some people, however, have extensive psoriasis or experience serious effects from psoriasis. Psoriatic arthritis produces symptoms very similar to those of rheumatoid arthritis. Rarely, psoriasis covers the entire body and produces exfoliative psoriatic dermatitis, in whic h

the entire skin becomes inflamed. This form of psoriasis is serious because, like a burn, it keeps the skin from serving as a protective barrier against injury and infection. In another uncommon form of psoriasis, pustular psoriasis, large and small pus-filled pimples (pustules) form on the palms of the hands and soles of the feet. Sometimes, these pustules are scattered on the body.

Psoriasis may flare up for no apparent reason, or a flare-up may result from severe sunburn, skin irritation, antimalaria drugs, lithium, beta-blocker drugs (such as propranolol and metoprolol), or almost any medicated ointment or cream. Streptococcal infections (especially in children), bruises, and scratches can also stimulate the formation of new plaques.

1.2. Molecular basis

Psoriasis is a chronic inflammatory disease. The inflammatory events in psoriasis are composed of a complex series of inductive and effector processes, which require the regulated expression of various proinflammatory genes. NF-κB is a protein transcription factor that is required for maximal transcription of many of these proinflammatory molecules. It consists of a heterodimer of the p50 and p65 proteins retained inactive in the cytoplasm tightly bound to the inhibitory subunit IκB. Upon activation, IκB is rapidly and sequentially phosphorylated by the action of IκB kinases, ubiquitinated, and degraded by the ubiquitin-proteasome. The active subunit (p50 and p65) is translocated to the nucleus, where it binds to cognate DNA sequences and stimulates gene transcription of proinflammatory genes.

1.3. Diagnosis

Psoriasis may be misdiagnosed at first because many other disorders can produce similar plaques and flaking. To confirm a diagnosis, a doctor may perform a skin biopsy by removing a skin specimen and examining it under a microscope.

1.4. Treatment

When a person has only a few small plaques, using ointments and creams that lubricate the skin (emollients) once or twice a day can keep the skin moist. Ointments containing corticosteroids, Vitamin D cream, salicylic acid or coal tar are effective in many patients with limited psoriasis. Stronger medications like anthralin are used sometimes, but they can irritate the skin and stain sheets and clothing. When the scalp is affected, shampoos containing these active ingredients are often used. For pustular psoriasis, the two most effective medications are etretinate and isotretinoin, which are also used to treat severe acne. Ultraviolet light also can help clear up psoriasis. In fact, during summer months, exposed regions of affected skin may clear up spontaneously. Sunbathing often helps to clear up the plaques on larger areas of the body; exposure to ultraviolet light under controlled conditions is another common therapy.

No drug for treating severe forms of psoriasis without severe side effects is marketed yet. For extensive psoriasis, ultraviolet therapy may be supplemented by psoralens, drugs that make the skin extra sensitive to the effects of ultraviolet light. The combination of psoralens and ultraviolet light (PUVA) is usually effective and may clear up the skin for several months. However, PUVA treatment can increase the risk of skin cancer from ultraviolet light; therefore, the treatment must be closely supervised by a doctor. For most serious forms of psoriasis and widespread psoriasis, a doctor may give methotrexate. Used to treat some forms of cancer, this drug interferes with the growth and multiplication of skin cells. It can be effective in extreme cases but may cause adverse effects on the bone marrow, kidneys, and liver. Another effective medication, cyclosporine, also has serious side effects.

New generation drugs that are currently under development include Efalizumab (Raptiva®), an humanized anti-CD11a antibody. It has been shown that Efalizumab, given subcutaneously once-weekly, provides clinical benefit in patients with moderate-to-severe plaque psoriasis (Cather et al. (2003) Expert Opin Biol Ther. 3:361-370). Efalizumab offers an new therapeutic option for the treatment of psoriasis and the potential for improved and potentially safer long-term, continuous "maintenance" therapy.

2. Psoriasis susceptibility loci

The multifactorial etiology of psoriasis is well established. Although environmental factors, such as streptococcal infections, affect the onset of the disease, family studies indicate a strong genetic component. Twin studies show the concordance in monozygotic twins to be 65 to 70% (Farber et al., 1974), compared to 15 to 20% in dizygotic twins. Family studies estimate the risk to first-degree relatives at between 8 to 23%. However, there are also several known environmental factors, including streptococcal infection and stress, that affect the onset and presentation of the disease.

Several psoriasis susceptibility loci have been mapped: PSORS1 on 6p21, PSORS2 on 17q, PSORS3 on 4q, PSORS4 on 1cen-q21, PSORS5 on 3q21, PSORS6 on 19p, PSORS7 on 1p, and PSORS8 on 4q31. The loci on 6p and 17q appear to be well established. Additional putative psoriasis candidate loci have been reported on 16q and 20p.

The major susceptibility locus for psoriasis is PSORS1 (Nair et al. 1997; Trembath et al. 1997; Oka et al. 1999; Lee et al. 2000; Veal et al. 2001). Several positional candidate genes are located within the PSORS1 susceptibility locus for psoriasis: HLA-C (the leukocyte antigen C), HCR (the α-helix-coiled-coil-rod homologue), POU5F1 (the octamer transcription factor 3), TCF19 (the cell growth-regulated gene), the corneodesmosin gene, a gene encoding a plectin-like protein and three genes displaying no homology to any known sequences in any DNA database.

Veal et al. performed a SNP-haplotype-based association analysis of PSORS1 to refine the susceptibility locus (Veal et al. 2002). They identified a 10-kb major region for susceptibility for psoriasis. They showed that this restricted region comprised two biallelic markers, SNPs n.7 and n.9, with probability values clearly exceeding any other markers studied before. This 10-kb region did not contain any known gene. In addition, database analysis of this restricted region did not allow the identification of any expressed gene, although a non-expressed pseudogene was identified. Since SNPs n.7 and n.9 lie in a non coding region, respectively 7 and 4 kb centromeric to HLA-C, Veal et al. concluded that SNPs n.7 and n.9 may lie within a regulatory region influencing expression of HLA-C.

3. The ubiquitin-proteasome pathway

The ubiquitin proteasome pathway has a central role in the selective degradation of intracellular proteins. Among the key proteins modulated by the proteasome are those involved in the control of inflammatory processes, cell cycle regulation, cell growth and gene expression. The proteasome is a large multimeric protease present in all eukaryotic cells that exhibits a highly conserved 20S core structure. Proteasomes are responsible for the degradation of protein substrates after they have been "tagged" by a poly-ubiquitin chain. Among others, the proteasome is known to be responsible for the degradation of IκB (Regnier et al. 1997. Cell. 90:373-383). Thus proteasome inhibition inhibits NFκB activation by blocking the degradation of its inhibitory protein IκB, and inhibition of the proteasome has been proposed as a potential mean to treat T cell-mediated disorders such as psoriasis (Zollner et al. 2002 J Clin Invest. 109:671-9).

The selective degradation of proteins through the ubiquitin proteasome pathway involves the activation of a signaling cascade that generates the covalent attachment of a polyubiquitin chain to protein targets. The polyubiquitin chain formed through the addition of multiple ubiquitin molecules to the target acts as a signal for degradation by the proteasome, a large multimeric protein complex. Ubiquitin conjugation requires the presence of three key enzymes: the ubiquitin-activating enzyme E1, the ubiquitin-conjugating enzyme E2 and the ubiquitin ligase E3.

De-ubiquitinating activities can promote the accumulation of ubiquitin in a given cell and are also thought to counteract the effects of E2/E3-mediated conjugation by removing the polyubiquitin chain from conjugated proteins prior to their degradation by the proteasome. This might either represent a means of preventing degradation by the proteasome, or might be part of those ubiquitination processes not aimed at directing protein degradation. De-ubiquitinating enzymes can be subdivided into two broad groups: ubiquitin C-terminal hydrolases (UCHs) and ubiquitin isopeptidases (UBPs) (Wilkinson, 1997). As far as UBPs are concerned, a number of published reports indicate that certain UBPs have highly specific functions. IsoT, a member of this family that has been studied in detail, is able to cleave both linear and isopeptide -linked

ubiquitin, and it appears to require a free ubiquitin C-terminus for optimal activity (Wilkinson et al. 1995).

Naviglio et al cloned and characterized the UBP8 ubiquitin isopeptidase in 1998 (EMBO J. 17:3241-3250). The biochemical activity of UBP8 was determined, and it was shown that UBP8 can both hydrolyze ubiquitin-isopeptide bonds and cleave purified linear ubiquitin chains. Down-regulation of UBP8 generates a substantial derangement of the overall cell protein ubiquitination, thus indicating that UBP8 plays a general role in the ubiquitin pathway. Moreover, microinjection of antisense UBP8 cDNA in quiescent human cells prevents S-phase entry, and microinjection of antisense UBP8 cDNA in growing osteosarcoma cells determines instead an accumulation of cells in S-phase. Thus Naviglio et al. showed that inhibition of the cellular ubiquitin isopeptidase UBP8 has a striking effect on cell proliferation. In 2000, UBP8 was shown to associate with Hrs-binding protein both *in vitro* and in cultured cells (Kato et al. J Biol Chem. 275:37481-37487). Hrs-binding protein together with Hrs plays a regulatory role in endocytic trafficking of growth factor-receptor complexes through early endosomes. Kato et al. hypothesized that UBP8 associated with Hbp plays a positive regulatory role in proteasomal and/or lysosomal degradation of growth factor receptors.

Accordingly, proteins of the ubiquitin-proteasome pathway have been shown to play an important role in, e.g., cell cycle regulation, regulation of cell proliferation and degradation of proteins involved in inflammation. Consequently, modulation of proteins of the ubiquitin-proteasome pathway is a treatment option for cancer and chronic inflammatory diseases such as, e.g., rheumatoid arthritis, asthma, inflammatory bowel disease, multiple sclerosis and psoriasis.

SUMMARY OF THE INVENTION

The present invention stems from the finding of an expressed gene located at human chromosome 6p21, within the 10-kb region that defines the major susceptibility locus for psoriasis. This gene, the UBP8rp gene, encodes a protein of the ubiquitin protea some pathway. The UBP8rp gene comprises two introns located at nucleotide positions 1018 to 1046 of SEQ ID NO: 1 and 1676 to 1718 of SEQ ID NO: 1.

Therefore, an aspect of the present invention relates to an isolated UBP8rp gene comprising introns having a sequence of (i) nucleotides 1018 to 1046 of SEQ ID NO: 1; and (ii) nucleotides 1676 to 1718 of SEQ ID NO: 1.

The present invention further relates to an isolated UBP8rp polynucleotide complementary to a messenger RNA transcribed from the UBP8rp gene.

The present invention further pertains to a purified UBP8rp polypeptide encoded by the UBP8rp gene or by a UBP8rp polynucleotide.

The present invention is further directed to an expression vector comprising the UBP8rp gene or a UBP8rp polynucleotide.

A host cell comprising the above expression vector is a further aspect of the present invention.

The present invention is further directed to a method of making a UBP8rp polypeptide, said method comprising the steps of culturing a host cell according to the invention under conditions suitable for the production of a UBP8rp polypeptide within said host cell.

A further aspect of the invention relates to an antibody that specifically binds to a UBP8rp polypeptide.

The use of a UBP8rp polypeptide as a target for screening for natural binding partners, the use of a UBP8rp polypeptide as a target for screening candidate modulators, and the use of a modulator of a UBP8rp polypeptide for preparing a medicament for the treatment of a chronic inflammatory disease are also within the present invention.

Further, the present invention pertains to a method of assessing the efficiency of a modulator of a UBP8rp polypeptide for the treatment of psoriasis, said method comprising administering said modulator to an animal model for psoriasis; wherein a determination that said modulator ameliorates a representative characteristic of psoriasis in said animal model indicates that said modulator is a drug for the treatment of psoriasis.

The present invention is further based on the finding of novel UBP8rp-related biallelic markers located within the major susceptibility locus for psoriasis. These UBP8rp-related biallelic markers are depicted in the table below:

Biallelic maker No. 1	Position on SEQ ID NO: 1	Alternative nucleotides
1	1199	A/G
2	1262	C/T
4	1444	G/T
6	1490	A/G
7	1505	G/T
10	1630	A/G
12	1680	A/G
13	1895	A/G
14	2180	A/G
15	2449	A/T
16	2721	G/T
17	3127	A/G
18	3137	C/T
19	3138	A/G
21	3222	C/G
22	3269	C/T

23	3445	C/T
24	3470	A/G
25	3915	C/T
26	3973	A/C
27	4254	A/G
28	4472	A/T
29	4660	C/T
31	4919	A/G
32	4973	C/T
33	5063	C/T
34	5065	G/T
35	5079	C/T
37	5088	C/G
38	5090	C/T
39	5407	C/T
40	5466	A/G
41	5520	C/T

Therefore, in a further aspect, the present invention is directed to the use of at least one UBP8rp-related biallelic marker for determining whether there is a significant association between said biallelic marker and a chronic inflammatory disease.

The present invention further relates to the use of at least one UBP8rp-related biallelic marker for diagnosing whether an individual suffers from or is at risk of suffering from a chronic inflammatory disease.

The invention also concerns a method of genotyping comprising the steps of: (a) isolating a nucleic acid from a biological sample; and (b) detecting the nucleotide present at one or more of the UBP8rp-related biallelic markers.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A to 1K show the annotation of the gene encoding UBP8rp.

Figure 2 shows an alignment between UBP8rp (SEQ ID NO: 3) and UBP8 (SEQ ID NO: 4).

Figure 3 shows the rhodanese domain of UBP8rp (SEQ ID NO: 3)

BRIEF DESCRIPTION OF THE SEQUENCES OF THE SEQUENCE LISTING

SEQ ID NO: 1 corresponds to the genomic region comprising the UBP8rp gene.

SEQ ID NO: 2 corresponds to the CDS coding for UBP8rp.

SEQ ID NO: 3 corresponds to the protein sequence of UBP8rp.

SEQ ID NO: 4 corresponds to the protein sequence of UBP8.

SEQ ID Nos. 5-51 correspond to primers.

DETAILED DESCRIPTION OF THE INVENTION

The present invention stems from the finding of an expressed gene located at human chromosome 6p21, within the 10-kb region that defines the major susceptibility locus for psoriasis. This gene codes for a novel protein of the ubiquitin-proteasome pathway, UBP8rp. Novel biallelic markers located in the UBP8rp gene are further provided.

Accordingly, the present invention provides novel UBP8rp polypeptides and means to identify compounds useful in the treatment of psoriasis and other chronic inflammatory diseases such as, e.g., psoriatic arthritis, rheumatoid arthritis, asthma, inflammatory bowel disease and multiple sclerosis. Specifically, the invention relates to the use of UBP8rp polypeptides as targets for screening for modulators thereof. The use of said modulators for treating psoriasis and other chronic inflammatory diseases, and the use of novel biallelic markers located in the UBP8rp gene for diagnosing psoriasis and other chronic inflammatory diseases are further aspects of the present invention.

1. Polynucleotides of the present invention

A first aspect of the present invention relates to an isolated gene comprising introns having a sequence of (i) nucleotides 1018 to 1046 of SEQ ID NO: 1; and (ii) nucleotides 1676 to 1718 of SEQ ID NO: 1.

As used herein, the term "intron" refers to a sequence of nucleotides interrupting the protein-coding sequences of a gene. Introns are transcribed into primary RNA but are cut out of the primary RNA to generate a messenger RNA that it is translated into protein.

As used herein, the term "gene" refers to a sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific protein. A gene usually comprises exons, introns, 5' and 3' untranslated regions, and upstream and downstream regulatory sequences. A gene may encode different isoforms of the same protein. These isoforms may be generated by, e.g., alternative splicing events or start of translation from alternative initiation codons. The term "gene", as used herein, does not include pseudogenes.

As further used in this specification, the term "UBP8-rp gene" refers to the gene comprising the introns shown at nucleotides 1018 to 1046 and nucleotides 1676 to 1718 of SEQ ID NO: 1. This gene is located at locus 6p21, within the major susceptibility locus for psoriasis, and codes for the UBP8rp protein.

The terms "comprising", "consisting of", or "consisting essentially of" have distinct meanings. However, each term may be substituted for another herein to change the scope of the invention.

The allelic variants of the UBP8rp gene are encompassed within the scope the present invention. Several alleles of the UBP8rp gene are shown on SEQ ID NO: 1. Furthermore,

procedures known in the art can be used to obtain other allelic variants of the UBP8rp gene using information from the sequences disclosed herein. For example, other allelic variants may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for allelic variants using any technique known to those skilled in the art.

Another aspect of the present invention relates to an isolated polynucleotide complementary to a messenger RNA transcribed from the gene of claim 1.

As further used herein, the term "UBP8rp polynucleotide" refers to an isolated polynucleotide complementary to a messenger RNA transcribed from the UBP8rp gene, or to a fragment thereof.

As used herein, the term "messenger RNA" (mRNA) refers to the processed RNA molecule that does not comprise any intron sequence. The term messenger RNA encompasses all alternative splice variants translated from the UBP8rp gene.

Such a messenger RNA may comprise any combination of exon of the UBP8rp gene. In one embodiment, the UBP8rp polynucleotide comprises exon 1 comprising nucleotides 1 to 167 of SEQ ID NO: 2. In another embodiment, the UBP8rp polynucleotide comprises exon 2 comprising nucleotides 168 to 796 of SEQ ID NO: 2. In another embodiment, the UBP8rp polynucleotide comprises exon 3 comprising nucleotides 797 to 1449 of SEQ ID NO: 2.

In a preferred embodiment, the UBP8rp polynucleotide comprises SEQ ID NO: 2.

Any procedures known in the art can be used to obtain UBP8rp polynucleotides. UBP8rp polynucleotides can for example be obtained as described in Example 1.

The present invention also encompasses UBP8rp polynucleotides for use as primers and probes. Such primers are useful in order to detect the presence of at least a copy of a UBP8rp polynucleotide, complement, or variant thereof in a test sample. The probes of the present invention are useful for a number of purposes. They can preferably be used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in the UBP8rp using other techniques. They may further be used for *in situ* hybridization. Preferred primers of the present invention are those of SEQ ID Nos. 5-51.

The present invention also encompasses polynucleotides UBP8rp polynucleotide that codes for a fragment of a UBP8rp polypeptide. The fragment may for example consist of an antigenic epitope of the UBP8rp and find use in production of antibodies.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid substrate, such as, e.g., a microarray. A substrate comprising a plurality of oligonucleotide primers or probes of the invention may be used either for detecting or amplifying targeted sequences in the UBP8rp gene, may be used for detecting

mutations in the coding or in the non-coding sequences of the UBP8rp mRNAs, and may also be used to determine expression of UBP8rp mRNAs in different contexts such as in different tissues, at different stages of a process (embryo development, disease treatment), and in patients versus healthy individuals.

2. Polypeptides of the present invention

Another aspect of the present invention relates to a purified polypeptide encoded by the UBP8rp gene or by a UBP8rp polynucleotide.

Preferably, the UBP8rp polypeptide is selected from the group consisting of:

a) a polypeptide comprising SEQ ID NO:3;
b) a polypeptide comprising a span of at least 470 amino acids of SEQ ID NO: 3;
c) a polypeptide comprising a span of at least 15 amino acids of SEQ ID NO: 3, wherein said span falls within amino acids 467 to 482 of SEQ ID NO: 3;
d) a mutein of any of (a) to (c), wherein the amino acid sequence has at least 95%, 96%, 97%, 98% or 99% identity to at least one of the sequences in (a) to (c);
e) a mutein of any of (a) to (c) which is encoded by a polynucleotide which hybridizes to the complement of a DNA sequence encoding any of (a) to (c) under highly stringent conditions; and
f) a mutein of any of (a) to (c) wherein any changes in the amino acid sequence are conservative amino acid substitutions to the amino acid sequences in (a) to (c).

The term "UBP8rp polypeptide" is used herein to embrace all of the polypeptides of the present invention.

In a preferred embodiment, the UBP8rp polypeptide corresponds to a full-length UBP8rp protein. The UBP8rp protein is a member of the ubiquitin proteasome pathway, as described in Example 1. UBP8rp plays a role in the ubiquitin-conjugation and de-ubiquitination of intracellular proteins, either by de-ubiquitinating said intracellular proteins, or by regulating ubiquitinating and de-ubiquitinating enzymes. The biological activity of a UBP8rp polypeptide refers to the modulation of the ubiquitination state of intracellular proteins by UBP8rp.

The present invention is also directed to polypeptides consisting of a fragment of at least 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470 or 480 amino acids of SEQ ID NO: 3. Preferably, said fragment falls within amino acids 467 to 482 of SEQ ID NO: 3.

The present invention is also directed to naturally occuring, recombinant, or chimeric polypeptides comprising any of the above fragments.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position

107 of SEQ ID NO: 3 is an arginine. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 107 of SEQ ID NO: 3 is a lysine.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 128 of SEQ ID NO: 3 is an threonine. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 128 of SEQ ID NO: 3 is a methionine.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 183 of SEQ ID NO: 3 is an asparagine. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 183 of SEQ ID NO: 3 is an histidine.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 189 of SEQ ID NO: 3 is an asparagine. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 189 of SEQ ID NO: 3 is a tyrosine.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 203 of SEQ ID NO: 3 is glycine. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 203 of SEQ ID NO: 3 is a glutamic acid.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 204 of SEQ ID NO: 3 is an arginine. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 204 of SEQ ID NO: 3 is a lysine.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 209 of SEQ ID NO: 3 is a glycine. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 209 of SEQ ID NO: 3 is a valine.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 251 of SEQ ID NO: 3 is an glycine. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 251 of SEQ ID NO: 3 is an arginine.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 325 of SEQ ID NO: 3 is an glutamic acid. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 325 of SEQ ID NO: 3 is a lysine.

One embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 420 of SEQ ID NO: 3 is an alanine. Another embodiment is directed to a UBP8rp polypeptide wherein the amino acid at position 420 of SEQ ID NO: 3 is a threonine.

Further embodiments are directed to muteins. As used herein the term "muteins" refers to analogs of UBP8rp, in which one or more of the amino acid residues of a natural UBP8rp are replaced by different amino acid residues, or are deleted, or one or more amino acid residues are added to the natural sequence of UBP8rp, without lowering considerably the activity of the resulting products as compared with the wild-type UBP8rp. These muteins are prepared by known synthesis and/or by site-directed mutagenesis techniques, or any other known technique suitable therefore.

Muteins of UBP8rp, which can be used in accordance with the present invention, or nucleic acid coding thereof, include a finite set of substantially corresponding sequences as substitution peptides or polynucleotides which can be routinely obtained by one of ordinary skill in the art, without undue experimentation, based on the teachings and guidance presented herein.

UBP8rp polypeptides in accordance with the present invention include proteins encoded by a nucleic acid, such as DNA or RNA, which hybridizes to DNA or RNA, which encodes UBP8RPb, in accordance with the present invention, under moderately or highly stringent conditions. The term "stringent conditions" refers to hybridization and subsequent washing conditions, which those of ordinary skill in the art conventionally refer to as "stringent". See Ausubel et al., Current Protocols in Molecular Biology, supra, Interscience, N.Y., §6.3 and 6.4 (1987, 1992), and Sambrook et al. (Sambrook, J. C., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Without limitation, examples of stringent conditions include washing conditions 12-20°C below the calculated Tm of the hybrid under study in, *e.g.*, 2 x SSC and 0.5% SDS for 5 minutes, 2 x SSC and 0.1% SDS for 15 minutes; 0.1 x SSC and 0.5% SDS at 37 °C for 30-60 minutes and then, a 0.1 x SSC and 0.5% SDS at 68°C for 30-60 minutes. Those of ordinary skill in this art understand that stringency conditions also depend on the length of the DNA sequences, oligonucleotide probes (such as 10-40 bases) or mixed oligonucleotide probes. If mixed probes are used, it is preferable to use tetramethyl ammonium chloride (TMAC) instead of SSC.

The polypeptides of the present invention include muteins having an amino acid sequence at least 50% identical, more preferably at least 60% identical, and still more preferably 70%, 80%, 90%, 95%, 96%, 97%, 98% or 99% identical to a UBP8RPb polypeptide of the present invention. By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% (5 of 100) of the amino acid residues in the subject sequence may be inserted, deleted, or substituted with another amino acid.

For sequences where there is not an exact correspondence, a "% identity" may be determined. In general, the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include inserting "gaps" in either one or both sequences, to enhance the degree of alignment. A % identity may be determined over the

whole length of each of the sequences being compared (so-called global alignment), that i particularly suitable for sequences of the same or very similar length, or over shorter, define lengths (so-called local alignment), that is more suitable for sequences of unequal length. Methods for comparing the identity and homology of two or more sequences are well known in the art. Thus for instance, programs available in the Wisconsin Sequence Analysis Package, version 9.1 (Devereux J et al. (1984) Nucleic Acids Res. 12:387-395), for example the program BESTFIT and GAP, may be used to determine the % identity between two poly nucleotides and the % identity and the % homology between two polypeptide sequences.. BESTFIT uses the "local homology" algorithm of Smith and Waterman (1981, J Mol Evol. 18:38-46) and finds the best single region of similarity between two sequences. Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs (Altschul et al. (1990) J Mol Biol. 215:403-410), accessible through the home page of the NCBI at world wide web site ncbi.nlm.nih.gov) and FASTA (Pearson (1990) Methods in Enzymology, 183:63-99; Pearson and Lipman (1988) Proc Nat Acad Sci USA, 85:2444-2448).

Preferred changes for muteins in accordance with the present invention are what are known as "conservative" substitutions. Conservative amino acid substitutions of UBP8rp polypeptides, may include synonymous amino acids within a group which have sufficiently similar physicochemical properties that substitution between members of the group will preserve the biological function of the molecule (Grantham (1974) Science 185:862-864). It is clear tha insertions and deletions of amino acids may also be made in the above-defined sequence without altering their function, particularly if the insertions or deletions only invo lve a few amino acids, e.g. under thirty, and preferably under ten, and do not remove or displace amino acids which are critical to a functional conformation, e.g. cysteine residues. Proteins and muteins produced by such deletions and/or insertions come within the purview of the present invention.

Preferably, the synonymous amino acid groups are those defined in Table I. More preferably, the synonymous amino acid groups are those defined in Table II; and most preferably the synonymous amino acid groups are those defined in Table III.

TABLE I

Preferred Groups of Synonymous Amino Acids

Amino Acid	Synonymous Group
Ser	Ser, Thr, Gly, Asn
Arg	Arg, Gln, Lys, Glu, His
Leu	Ile, Phe, Tyr, Met, Val, Leu
Pro	Gly, Ala, Thr, Pro
Thr	Pro, Ser, Ala, Gly, His, Gln, Thr
Ala	Gly, Thr, Pro, Ala
Val	Met, Tyr, Phe, Ile, Leu, Val
Gly	Ala, Thr, Pro, Ser, Gly
Ile	Met, Tyr, Phe, Val, Leu, Ile

Phe	Trp, Met, Tyr, Ile, Val, Leu, Phe
Tyr	Trp, Met, Phe, Ile, Val, Leu, Tyr
Cys	Ser, Thr, Cys
His	Glu, Lys, Gln, Thr, Arg, His
Gln	Glu, Lys, Asn, His, Thr, Arg, Gln
Asn	Gln, Asp, Ser, Asn
Lys	Glu, Gln, His, Arg, Lys
Asp	Glu, Asn, Asp
Glu	Asp, Lys, Asn, Gln, His, Arg, Glu
Met	Phe, Ile, Val, Leu, Met
Trp	Trp

TABLE II

More Preferred Groups of Synonymous Amino Acids

Amino Acid	Synonymous Group
Ser	Ser
Arg	His, Lys, Arg
Leu	Leu, Ile, Phe, Met
Pro	Ala, Pro
Thr	Thr
Ala	Pro, Ala
Val	Val, Met, Ile
Gly	Gly
Ile	Ile, Met, Phe, Val, Leu
Phe	Met, Tyr, Ile, Leu, Phe
Tyr	Phe, Tyr
Cys	Cys, Ser
His	His, Gln, Arg
Gln	Glu, Gln, His
Asn	Asp, Asn
Lys	Lys, Arg
Asp	Asp, Asn
Glu	Glu, Gln
Met	Met, Phe, Ile, Val, Leu
Trp	Trp

TABLE III

Most Preferred Groups of Synonymous Amino Acids

Amino Acid	Synonymous Group
Ser	Ser
Arg	Arg
Leu	Leu, Ile, Met
Pro	Pro
Thr	Thr
Ala	Ala
Val	Val
Gly	Gly
Ile	Ile, Met, Leu
Phe	Phe
Tyr	Tyr
Cys	Cys, Ser
His	His
Gln	Gln

Asn	Asn
Lys	Lys
Asp	Asp
Glu	Glu
Met	Met, Ile, Leu
Trp	Met

Examples of production of amino acid substitutions in proteins which can be used for obtaining muteins of UBP8rp, polypeptides for use in the present invention include any known method steps, such as presented in US patents 4,959,314, 4,588,585 and 4,737,462, to Mark et al; 5,116,943 to Koths et al., 4,965,195 to Namen et al; 4,879,111 to Chong et al; and 5,017,691 to Lee et al; and lysine substituted proteins presented in US patent No. 4,904,584 (Shaw et al).

Preferably, the muteins of the present invention exhibit substancially the same biological activity as the UBP8RPb polypeptide to which it corresponds.

In other embodiments, UBP8rp polypeptides do not exhibit the biological activity as the UBP8RPb polypeptide to which it corresponds. Other uses of the polypeptides of the present invention include, *inter alia,* as epitope tags, in epitope mapping, and as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods known to those of skill in the art. Such polypeptides can be used to raise polyclonal and monocl onal antibodies, which are useful in assays for detecting UBP8rp expression, or for purifying UBP8rp. As a matter of example, a further specific use for UBP8rp polypeptides is the use of such polypeptides the yeast two-hybrid system to capture UBP8rp binding proteins, which are candidate modulators according to the present invention, as further detailed below.

3. Vectors, host cells and host organisms of the present invention

The present invention also relates to vectors comprising the UBP8rp gene or a UBP8rp polynucleotide. More particularly, the present invention relates to expression vectors which include the UBP8rp gene or a UBP8rp polynucleotide. Preferably, such expression vectors comprise a polynucleotide encoding a UBP8rp polypeptide.

The term "vector" is used herein to designate either a circular or a linear DNA or RNA compound, which is either double-stranded or single-stranded, and which comprise at least one polynucleotide of the present invention to be transferred in a cell host or in a unicell ular or multicellular host organism. An "expression vector" comprises appropriate signals in the vectors, said signals including various regulatory elements, such as enhancers/promoters from both viral and mammalian sources that drive expression of the ins erted polynucleotide in host cells. Selectable markers for establishing permanent, stable cell clones expressing the products such as, e.g., a dominant drug selection, are generally included in the expression vectors of the

invention, as they are elements that link expression of the drug selection markers to expression of the polypeptide.

Additionally, the expression vector may be a fusion vector driving the expression of a fusion polypeptide between a UBP8rp polypeptide and a heterologous polypeptide. For example, the heterologous polypeptide may be a selectable marker such as, e.g, a luminescent protein, or a polypeptide allowing the purification of the fusion polypeptide.

The polynucleotides of the present invention may be used to, e.g., express the encoded polypeptide in a host cell for producing the encoded polypeptide. The polynucleotides of the present invention may further be used to express the encoded polypeptide in a host cell for screening assays. Screening assays are of particular interest for identifying modulators and/or binding partners of UBP8rp polypeptides as further detailed below. The polynucleotides of the present invention may also be used to express the encoded polypeptide in a host organism for producing a beneficial effect. In such procedures, the encoded protein may be transiently expressed in the host organism or stably expressed in the host organism. The encoded polypeptide may have any of the properties described herein. The encoded polypeptide may be a protein which the host organism lacks or, alternatively, the encoded protein may augment the existing levels of the protein in the host organism.

In one embodiment, the expression vector is a gene therapy vector. Viral vector systems that have application in gene therapy have been derived from, e.g., adenoviral vectors and retroviral vectors.

Another object of the invention comprises a host cell comprising the UBP8rp gene or a UBP8rp polynucleotide. Such host cells may have been transformed, transfected or transduced with a polynucleotide encoding a UBP8rp polypeptide. Also included are host cells that are transformed, transfected or transduced with a recombinant vector such as one of those described above. The cell hosts of the present invention can comprise any of the polynucleotides of the present invention.

Any host cell known by one of skill in the art may be used. Preferred host cells used as recipients for the polynucleotides and expression vectors of the invention include:

a) Prokaryotic host cells: *Escherichia coli* strains (I.E.DH5-α strain), *Bacillus subtilis*, *Salmonella typhimurium*, and strains from species like *Pseudomonas*, *Streptomyces* and *Staphylococcus*.

b) Eukaryotic host cells: CHO (ATCC No. CCL-61), HeLa cells (ATCC No.CCL2; No.CCL2.1; No.CCL2.2), Cv 1 cells (ATCC No.CCL70), COS cells (ATCC No.CRL1650; No.CRL1651), Sf-9 cells (ATCC No.CRL1711), C127 cells (ATCC No. CRL-1804), 3T3 (ATCC No. CRL-6361), human kidney 293. (ATCC No. 45504; No. CRL-1573), BHK (ECACC No. 84100501; No. 84111301), *Saccharomyces cerevisiae* strains such as AH109 and Y184, and *Aspergillus niger* strains.

Another object of the invention comprises methods of making the above vectors and host cells by recombinant techniques. Any well-known technique for constructing an expression vector and for delivering it to a cell may be used for construction and delivering the vectors of the present invention. Such techniques include but are not limited to the techniques detailed in the examples.

Another object of the present invention is a transgenic animal which includes within a plurality of its cells a cloned recombinant UBP8rp polynucleotide. The terms "transgenic animals" or "host animals" are used herein to designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. The cells affected may be somatic, germ cells, or both. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention. In one embodiment, the invention encompasses non-human host mammals and animals comprising a recombinant vector of the invention or a UBP8rp polynucleotide disrupted by homologous recombination with a knock out vector.

In a preferred embodiment, these transgenic animals may be good experimental models in order to study diverse pathologies related to UBP8rp function. In particular, a transgenic animal wherein (i) an antisense mRNA binding to naturally occurring UBP8rp mRNAs is transcribed; or (ii) an mRNA expressing a UBP8rp polypeptide; may be a good animal model for psoriasis and/or other chronic inflammatory diseases.

4. <u>Methods of making the polypeptides of the present invention</u>

The present invention also relates to methods of making a UBP8rp polypeptide.

In one embodiment, the UBP8rp polypeptides of the present invention are isolated from natural sources, including tissues and cells, whether directly isolated or cultured cells, of humans or non-human animals. Soluble forms of UBP8rp may be isolated from body fluids. Methods for extracting and purifying natural membrane spanning proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by, e.g., differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis. The method described in Example 4 may for example be used. Polypeptides of the invention also can be purified from natural sources using antibodies directed against the polypeptides of the invention, such as those described herein, in methods which are well known in the art of protein purification.

In a preferred embodiment, the UBP8rp polypeptides of the invention are recombinantly produced using routine expression methods known in the art. The polynucleotide encoding the

desired polypeptide is operably linked to a promoter into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems may be used in forming recombinant polypeptides. The polypeptide is then isolated from lysed cells or, if a solu ble form is produced, from the culture medium and purified to the extent needed for its intended use.

Consequently, a further embodiment of the present invention is a method of making a polypeptide of the present invention, said method comprising the ste ps of:

a) obtaining a polynucleotide encoding a UBP8rp polypeptide;
b) inserting said polynucleotide in an expression vector such that the polynucleotide is operably linked to a promoter; and
c) introducing said expression vector into a host cell whereby sai d host cell produces said polypeptide.

In a preferred embodiment, the method further comprises the step of isolating the polypeptide. The skilled person will appreciate that any step of this method may be carried out separately. The product of each step may be transferred to another step in order to carry out the subsequent step.

In further embodiments, said polynucleotide consists of a coding sequence. In another aspect of this embodiment, said polynucleotide is a polynucleotide comprising SEQ ID NO: 2 o r a fragment thereof.

A further aspect of the invention relates to a method of making a polypeptide, said method comprising the steps of culturing a host cell comprising an expression vector comprising a UBP8rp polynucleotide under conditions suitable for the production of a UBP8rp polypeptide within said host cell. In a preferred embodiment, the method further comprises the step of purifying said polypeptide from the culture.

In another embodiment, it is often advantageous to add to the recombinant polynucleotide encoding a UBP8rp polypeptide additional nucleotide sequence which codes for secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues or GST tags, or an additional sequence for stability during recombinant production. Soluble portions of the UBP8rp polypeptide may be, e.g., linked to an Ig-Fc part in order to generate stable soluble variants.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including but not limited to differential extraction, ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, high performance liquid chromatography, phosphocellulose chromatography, hydrophob ic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, immunochromatography and lectin chromatography.

The expressed UBP8rp polypeptide may be purified using any standard immunochromatography techniques. In such procedure s, a solution containing the polypeptide

of interest, such as the culture medium or a cell extract, is applied to a column having antibodies against the polypeptide attached to the chromatography matrix. The recombinant protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound secreted protein is then released from the column and recovered using standard techniques.

5. Antibodies of the present invention

The present invention further relates to antibodies that specifically bind to the UBP8rp polypeptides of the present invention. More specifically, said antibodies bind to the epitopes of the polypeptides of the present invention. The antibodies of the present invention include IgG (including IgG1, IgG2, IgG3, and IgG4), IgA (including IgA1 and IgA2), IgD, IgE, or IgM, and IgY. The term "antibody" (Ab) refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where a binding domain is formed from the folding of variable domains of an antibody compound to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen, which allows an immunological reaction with the antigen. As used herein, the term "antibody" is meant to include whole antibodies, including single-chain whole antibodies, and antigen binding fragments thereof. In a preferred embodiment the antibodies are human antigen binding antibody fragments of the present invention include, but are not limited to, Fab, Fab' F(ab)2 and F(ab')2, Fd, single-chain Fvs (scFv), single-chain antibodies, disulfide-linked Fvs (sdFv) and fragments comprising either a V_L or V_H domain. The antibodies may be from any animal origin including birds and mammals. Preferably, the antibodies are from human, mouse, rabbit, goat, guinea pig, camel, horse or chicken. The present invention further includes humanized monoclonal and polyclonal antibodies, which specifically bind the polypeptides of the present invention.

Preferred antibodies of the present invention recognize an epitope within amino acids 467 to 482 of SEQ ID NO: 3, wherein said one or more amino-acids are required for binding of the antibody to a UBP8rp polypeptide.

A preferred embodiment of the invention is a method of specifically binding an antibody of the present invention to a UBP8rp polypeptide. This method comprises the step of contacting the antibody of the present invention with a UBP8rp polypeptide under conditions in which said antibody can specifically bind to said polypeptide. Such conditions are well known to those skilled in the art. This method may be used to, e.g., detect, purify, or activate or inhibit the activity of UBP8rp polypeptides.

The invention further relates to antibodies that act as modulators of the polypeptides of the present invention. Preferred antibodies are modulators that enhance the binding activity or

the biological activity of the UBP8rp polypeptide to which they bind. These antibodies may act as modulators for the biological activity of the UBP8rp polypeptide.

6. Uses of the polypeptides of the present invention

The present invention is also directed to the use of a UBP8rp polypeptide as a target for screening candidate modulators.

As used herein, the term "modulator" refers to a compound that increases or decreases any of the properties of a UBP8rp polypeptide. As used herein, a "UBP8rp modulator" refers to a compound that increases or decreases the activity of a UBP8rp polypeptide and/or to a compound that increases or decreases the trasncription level of the UBP8rp mRNA. The term "modulator" encompasses both agonists and antagonists.

As used herein, a "UBP8rp antagonist" refers to a compound that decreases the activity of a UBP8rp polypeptide and/or to a compound that decreases the expression level of the UBP8rp mRNA encoding said polypeptide. The terms "antagonist" and "inhibitor" are considered to be synonymous and can be used interchangeably throughout the disclosure.

As used herein, a "UBP8rp agonist" refers to a compound that increases the activity of a UBP8rp polypeptide and/or to a compound that increases the expression level of the UBP8rp mRNA encoding said polypeptide. The terms "agonist" and "activator" are considered to be synonymous and can be used interchangeably throughout the disclosure.

Methods that can be used for testing modulators for their ability to increase or decrease the activity of a UBP8RP polypeptide or to increase or decrease the expression of a UBP8RP mRNA are well known in the art and further detailed below. These assays can be performed either *in vitro* or *in vivo*.

Candidate compounds according to the present invention include naturally occurring and synthetic compounds. Such compounds include, e.g., natural ligands, small molecules, antisense mRNAs, antibodies, aptamers and small interfering RNAs. As used herein, the term "natural ligand" refers to any signaling molecule that binds to a phosphatase comprising PP2A/Bγ *in vivo* and includes molecules such as, e.g., lipids, nucleotides, polynucleotides, amino acids, peptides, polypeptides, proteins, carbohydrates and inorganic molecules. As used herein, the term "small molecule" refers to an organic compound. As used herein, the term "antibody" refers to a protein produced by cells of the immune system or to a fragment thereof that binds to an antigen. As used herein, the term "antisense mRNA" refers an RNA molecule complementary to the strand normally processed into mRNA and translated, or complemantary to a region thereof. As used herein, the term "aptamer" refers to an artificial nucleic acid ligand (see, e.g., Ellington and Szostak (1990) Nature 346:818-822). As used herein, the term "small interfering RNA" refers to a double-stranded RNA inducing sequence-specific posttranscriptional gene silencing (see, e.g., Elbashir et al. (2001) Nature. 411:494-498).

Such candidate compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including, e.g., biological libraries, spatially addressable parallel solid phase or solution phase libraries, and synthetic library methods using affinity chromatography selection. The biological library approach is generally u sed with peptide libraries, while the other four approaches are applicable to peptide, non -peptide oligomers, aptamers or small molecule libraries of compounds.

One example of a method that may be used for screening candidate compounds for a modulator is a method comprising the steps of:

a) contacting a UBP8rp polypeptide with the candidate compound; and
b) testing the activity of said UBP8rp polypeptide in the presence of said candidate compound,

wherein a difference in the activity of said UBP8rp polypeptide in the presence of said compound in comparison to the activity in the absence of said compound indicates that the compound is a modulator of said UBP8rp polypeptide.

Alternatively, the assay may be a cell-based assay comprising the steps of:

a) contacting a cell expressing a UBP8rp polypeptide with the candidate compound; and
b) testing the activity of said UBP8rp polypeptide in the presence of said candidate compound,

wherein a difference in the activity of said UBP8rp polypeptide in the presence of said compound in comparison to the activity in the absence of said compound indicates that the compound is a modulator of said UBP8rp polypeptide.

The modulator may be an inhibitor or an activator. An inhibitor may decrease UBP8rp activity by, e.g., 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or 100% compared to UBP8rp activity in the absence of said inhibitor. An activator may increase UBP8rp activity by, e.g., 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or 100% compared to UBP8rp activity in the absence of said activator.

The modulator may modulate any activity of said UBP8RP polypeptide. The modulator may for example modulate UBP8rp mRNA expression within a cell, modulate the enzymatic activity of the UBP8rp polypeptide, or modulate binding of the UBP8rp polypeptide to its natural binding partners. Preferably, the activity of the UBP8RP polypeptide is assessed by measuring the ubiquitination state of proteins.

In one embodiment, the activity of a UBP8rp polypeptide is assessed by measuring the ubiquitin-conjugation and/or de-ubiquitination of proteins. Assays for measuring the ubiquitin-conjugation and/or de-ubiquitination of proteins are known by those of skill the art. Such assays

are described, e.g., by Naviglio et al. (1998, EMBO J. 17:3241-3250) and by Gnesutta et al. (2001, J Biol Chem. 276:39448-39454).

In a preferred embodiment, the activity of a UBP8rp polypeptide is assessed by measuring the de-ubiquitinating activity of said polypeptide. The de-ubiquitinating activity of a UBP8rp polypeptide may be measured by replacing UBP8 by a UBP8rp polypeptide in the de-ubiquitination assay described at page 3248 of Naviglio et al. (1998).

In another embodinent, the activity of a UBP8rp polypeptide is assessed by measuring the de-ubiquitinating activity of UBP8 in the presence of said UBP8rp polypeptide.

In a further preferred embodiment, the activity of a UBP8RP polypeptide is assessed by measuring the UBP8rp mRNA levels within a cell. In this embodiment, the activity can for example be measured using Northern blots, RT-PCR, quantitative RT-PCR with primers and probes specific for UBP8RP mRNAs. Alternatively, the expression of the UBP8RP mRNA is measured at the polypeptide level, by using labeled antibodies that specifically bind to the UBP8rp polypeptide in immunoassays such as ELISA assays, or RIA assays, Western blots or immunohistochemical assays.

Modulators of UBP8rp polypeptides, which may be found, e.g., by any of the above screenings, are candidate drugs for the treatment of a chronic inflammatory disease. Thus a preferred embodiment of the present invention is the use of a UBP8rp polypeptide as a target for screening candidate compounds for candidate drugs for the treatment of a chronic inflammatory disease.

As used herein, the term "<u>chronic inflammatory disease</u>" refers to a chronic pathologic inflammation of a tissue or an organ of an individual. Chronic inflammatory diseases include, e.g., psoriasis, psoriatic arthritis, rheumatoid arthritis, asthma, inflammatory bowel disease and multiple sclerosis. Preferably, said chronic inflammatory disease is psoriasis.

A further aspect of the present invention is the use of a modulator of a UBP8rp polypeptide for screening for drugs for the treatment of a chronic inflammatory disease. One example of a method that can be used for screening for drugs for the treatment of a chronic inflammatory disease and/or for assessing the efficiency of an modulator of a UBP8rp polypeptide for the treatment of a chronic inflammatory disease is a method comprising the st ep of administering said modulator to an animal model for said chronic inflammatory disease, wherein a determination that said modulator ameliorates a representative characteristic of said chronic inflammatory disease in said animal model indicates that sa id modulator is a drug for the treatment of said chronic inflammatory disease. Preferably, said chronic inflammatory disease is psoriasis.

Animal models for chronic inflammatory diseases and assays for determining whether a compound ameliorates a representative characteristic of the chronic inflammatory disease in

said animal model are known by those of skill in the art. A preferred animal model for psoriasi: is the SCID-hu mouse that is described in Zollner et al. (2002, J Clin Invest. 109:671-679).

Determining whether the modulator ameliorates a representative characteristic of : chronic inflammatory disease may be performed using several methods available in the ar Specifically, when studying psoriasis, the representative characteristic may be the Nationa Psoriasis Foundation Psoriasis Score (NPF-PS), the Psoriasis Area Severity Index scor (PASI), or Physician's Global Assessment score (PGA) (see, e.g., Gottlieb et al. (2003) J Drug Dermatol. 2:260-266).

In one preferred embodiment of the present invention, the representative characteristic i: the Psoriasis Area and Severity Index score. The Psoriasis Area and Severity Index is a measure of overall psoriasis severity and coverage (Fredriksson et al. (1978) Dermatologica 157:238-244). It is a commonly used measure in clinical trials for psoriasis treatments.

In a further embodiment, a determination that a modulator of a UBP8rp polypeptid ameliorates the PASI score of an animal model for psoriasis indicates that said modulator is a drug for the treatment of psoriasis. Preferably, a 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95% or greater improvement in PASI scores indicates that said modulator is a drug for the treatmen of psoriasis. Most preferably, a 75% or greater improvement in PASI scores (PASI 75) indicate that said modulator is a drug for the treatment of psoriasis.

A further aspect of the present invention is directed to the use of a modulator of a UBP8rp polypeptide for preparing a medicament for the treatment of a chronic inflammator disease. Such a medicament comprises said modulator of a UBP8rp polypeptide in combinatior with any physiologically acceptable carrier. Physiologically acceptable carriers can be preparec by any method known by those skilled in the art. Physiologically acceptable carriers include bu are not limited to those described in Remington's Pharmaceutical Sciences (Mack Publishing Company, Easton, USA 1985). Pharmaceutical compositions comprising a modulator of a UBP8RP polypeptide and a physiologically acceptable carrier can be for, e.g., intravenous topical, rectal, local, inhalant, subcutaneous, intradermal, intramuscular, oral, intracerebral anc intrathecal use. The compositions can be in liquid (e.g., solutions, suspensions), solid (e.g., pills, tablets, suppositories) or semisolid (e.g., creams, gels) form. Dosages to be administerec depend on individual needs, on the desired effect and the chosen route of administration.

Such a medicament comprising (i) a UBP8rp modulator; or (ii) a gene therapy vector of the invention may be used in combination with any known drug for the treatment of a chronic inflammatory disease. For example, when treating psoriasis, the modulator may be administered in combination with Raptiva, Tazarotene, Anapsos, Alefacept, Micanol, Efalith Olopatadine, Calcipotriol, Cyclosporin A, Halobetasol propionate, Halometasone, Acitretin, GMDP, Silkis, Betamethasone mousse, Clobetasol propionate foam, Tacalcitol and/or Falecalcitriol.

The present invention further relates to the use of a UBP8rp polypeptide for screening for natural binding partners. Using a UBP8rp polypeptide as a target has a great utility for the identification of proteins involved in psoriasis and for providing new intervention points in the treatment of chronic inflammatory diseases. Such methods for screening for natural binding partners of a UBP8rp polypeptide are well known in the art. One method for the screening of a candidate substance interacting with a UBP8rp polypeptide of the present invention comprises the following steps :

a) providing a polypeptide consisting of a UBP8rp polypeptide;

b) obtaining a candidate polypeptide;

c) bringing into contact said polypeptide with said candidate polypeptide;

d) detecting the complexes formed between said polypeptide and said candidate polypeptide.

In one embodiment of the screening method defined above, the complexes formed between the polypeptide and the candidate substance are further incubated in the presence of a polyclonal or a monoclonal antibody that specifically binds to the UBP8rp polypeptide.

In a particular embodiment of the screening method, the candidate is the expression product of a DNA insert contained in a phage vector (Parmley and Smith (1988) Gene. 73:305-318). Specifically, random peptide phage libraries are used. The random DNA inserts encode for polypeptides of 8 to 20 amino acids in length (see, e.g., Oldenburg et al. (1992) Proc Natl Acad Sci U S A. 89:5393-5397; Valadon et al. (1996) J Immunol Methods. 197:171-179). According to this particular embodiment, the recombinant phages expressing a polypeptide that binds to the immobilized UBP8rp polypeptide is retained and the complex formed between the UBP8rp polypeptide and the recombinant phage may be subsequently immunoprecipitated by a polyclonal or a monoclonal antibody directed against the UBP8rp polypeptide.

In a further particular embodiment of the screening method, the binding partners are identified through a two-hybrid screening assay. The yeast two-hybrid system is designed to study protein-protein interactions *in vivo* (Fields and Song (1989) Nature. 340:245-6), and relies upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein. This technique is also described in US Patent Nos. 5,667,973 and 5,283,173. The general procedure of library screening by the two-hybrid assay may for example be performed as described by Fromont-Racine et al. (1997, Nat Genet. 16:277-282), the bait polypeptide consisting of a UBP8rp polypeptide. More precisely, a UBP8rp polynucleotide is fused in frame to a polynucleotide encoding the DNA binding domain of the GAL4 protein, the fused nucleotide sequence being inserted in a suitable expression vector, for example pAS2 or pM3.

In a further particular embodiment of the screening method, the binding partners are identified through affinity chromatography. The UBP8rp polypeptide may be attached to the column using conventional techniques including chemical coupling to a suitable column matrix

(e.g. agarose, Affi Gel®, etc.). In some embodiments of this method, the affinity column contains chimeric proteins in which the UBP8rp polypeptide, or a fragment thereof, is fused to glutathion S transferase (GST). A mixture of cellular proteins or pool of expressed proteins as described above is applied to the affinity column. Polypeptides interacting with the UBP8rp polypeptide attached to the column can then be isolated and analyzed, e.g., on 2-D electrophoresis gel as described in Ramunsen *et al.*, (1997, Electrophoresis, 18:588-598). Alternatively, the proteins retained on the affinity column can be purified by electrophoresis-based methods and sequenced.

In a further particular embodiment of the screening method, the binding partners are identified through optical biosensor methods (see, e.g., Edwards and Leatherbarrow, 1997). This technique permits the detection of interactions between molecules in real time, without the need of labeled molecules.

7. <u>Biallelic markers of the present invention</u>

The present invention is directed to the use of at least one UBP8rp-related biallelic marker selected from the group consisting of the biallelic markers shown below for determining whether there is a significant association between said biallelic marker and a chronic inflammatory disease:

Biallelic maker No.	Position on SEQ ID NO: 1	Alternative nucleotides
1	1199	A/G
2	1262	C/T
4	1444	G/T
6	1490	A/G
7	1505	G/T
10	1630	A/G
12	1680	A/G
13	1895	A/G
14	2180	A/G
15	2449	A/T
16	2721	G/T
17	3127	A/G
18	3137	C/T
19	3138	A/G
21	3222	C/G
22	3269	C/T
23	3445	C/T
24	3470	A/G
25	3915	C/T
26	3973	A/C
27	4254	A/G
28	4472	A/T
29	4660	C/T
31	4919	A/G
32	4973	C/T
33	5063	C/T

34	5065	G/T
35	5079	C/T
37	5088	C/G
38	5090	C/T
39	5407	C/T
40	5466	A/G
41	5520	C/T

As used herein, the term "biallelic marker" refers to a polymorphism having two alleles at a fairly high frequency in the population, preferably a single nucleotide polymorphism. Typically the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater than 1%, preferably the frequency is greater than 10%, more preferably the frequency is at least 20% (i.e. heterozygosity rate of at least 0.32), even more preferably the frequency is at least 30% (i.e. heterozygosity rate of at least 0.42). In the present specification, the term "biallelic marker" is used to refer both to the polymorphism and to the locus carrying the polymorphism. As used herein, the term "UBP8rp-related biallelic marker" refers to a biallelic marker located in an exon of the UBP8rp gene, in an intron of the UBP8rp gene, or in regulatory regions of the UBP8rp gene. The term "UBP8rp-related biallelic marker of the present invention" refers to Biallelic markers 1, 2, 4, 6, 7, 10, 12-19, 21-30, 31-35 and 37-41 shown above and further described in Example 3.

Determining whether there is a significant association between said biallelic marker and a chronic inflammatory disease can be performed using any method well known by those of skill in the art. For example, the UBP8rp-related biallelic marker of the present invention may be genotyped in case and control populations for the inflammatory disease to be studied. The allelic frequency of markers between cases and controls may be investigated using, e.g., the Pearson Chi squared test. The EM (Expectation-Maximization) algorithm (Excoffier L & Slatkin M, 1995) may be used to estimate haplotypes for the population under investigation. Alternatively, haplotype frequency estimations may be performed by applying the OMNIBUS likelihood ratio test (PCT publication WO 01/091026). The association between UBP8rp-related biallelic markers of the present invention and psoriasis may also be performed as described by Veal et al (2002).

In all aspects and embodiments relating to UBP8rp-related biallelic markers of the present invention, the chronic inflammatory disease is preferably selected from the group consisting of psoriasis, psoriatic arthritis, rheumatoid arthritis, asthma, inflammatory bowel disease and multiple sclerosis. Most preferably, the chronic inflammatory disease is psoriasis.

The present invention is further directed to the use of at least one UBP8rp-related biallelic marker of the present invention for diagnosing whether an individual suffers from or is at risk of suffering from a chronic inflammatory disease.

In one embodiment, a single biallelic marker is used for diagnosing whether an individual suffers from or is at risk of suffering from a chronic inflammatory disease by determining the

genotype of an individual. In another embodiment, a combination of several biallelic markers may be used for diagnosing whether an individual suffers from or is at risk of suffering from a chronic inflammatory disease by determining the haplotype of an individual. For example, a two-markers haplotype, a three-markers haplotype or a four-markers haplotype may be determined.

As used herein, the term "genotype" refers to the identity of the alleles present in an individual or a sample. The term "genotype" preferably refers to the description of both copies of a single biallelic marker that are present in the genome of an individual. The individual is homozygous if the two alleles of the biallelic marker present in the genome are identical. The individual is heterozygous if the two alleles of the biallelic marker present in the genome are different.

The term "genotyping" a sample or an individual for a biallelic marker involves determining the specific alleles or the specific nucleotides carried by an individual at a biallelic marker.

As used herein, the term "haplotype" refers to a set of alleles of closely linked biallelic markers present on one chromosome and which tend to be inherited together.

Methods for determining the alleles, genotypes or haplotypes carried by an individual are well known by those of skill in the art and further detailed below.

In the context of the present invention, the individual is generally understood to be human.

UBP8rp-related biallelic markers 20 and 36 are highly associated with psoriasis, yielding p-values inferior to 10^{-9} (Veal et al., 2002). Thus a preferred embodiment of the present invention is directed to the use of (i) at least one UBP8rp-related biallelic marker of the present invention; and (ii) the biallelic marker 20 and/or the biallelic marker 36 for diagnosing whether an individual suffers from or is at risk of suffering from psoriasis.

The present invention is further directed to a method of genotyping comprising the steps of:

a) isolating a nucleic acid from a biological sample; and
b) detecting the nucleotide present at one or more of the UBP8rp-related biallelic markers of the present invention.

Preferably, said biological sample is derived from a single subject. It is preferred that the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome. In a preferred embodiment, the identity of the nucleotide at said biallelic marker is determined by a microsequencing assay. Preferably, a portion of a sequence comprising the biallelic marker is amplified prior to the determination of the identity of the nucleotide. The amplification may preferably be performed by PCR. Methods

of genotyping are well known by those of skill in the art and any other known protocol may be used. The nucleotide present at a UBP8rp-related biallelic marker of the present invention may for example be determined as described in Example 3.

Methods well-known to those skilled in the art that may be used for genotyping in order to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) (Orita et al. (1989) Proc Natl Acad Sci USA 86:2766-2770), denaturing gradient gel electrophoresis (DGGE) (Borresen et al. (1988) Mutat Res. 202:77-83.), heteroduplex analysis (Lessa et al. (1993) Mol Ecol. 2:119-129), mismatch cleavage detection (Grompe et al. (1989) Proc Natl Acad Sci USA. 86:5888-5892). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent No. 4,656,127. Oligonucleotide microarrays or solid-phase capturable dideoxynucleotides and mass spectrometry may also be used (Wen et al. (2003) World J Gastroenterol. 9:1342-1346; Kim et al. (2003) Anal Biochem. 316:251-258). Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, microsequencing assay, enzyme-based mismatch detection assay, or hybridization assay.

As used herein, the term "biological sample" refers to a sample comprising nucleic acids. Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like.

Methods of genotyping find use in, e.g., in genotyping case-control populations in association studies as well as in genotyping individuals in the context of detection of alleles of biallelic markers which are known to be associated with a given trait. In the context of the present invention, a preferred trait is a chronic inflammatory disease selected from the group of psoriasis, psoriatic arthritis, rheumatoid arthritis, asthma, inflammatory bowel disease and multiple sclerosis, and most preferably psoriasis.

In one embodiment, the above genotyping method further comprises the step of correlating the result of the genotyping steps with a risk of suffering from a chronic inflammatory disease.

The present invention is further directed to the use of at least one UBP8rp-related biallelic marker of the present invention for determining the haplotype of an individual. When determining the haplotype of an individual, each single chromosome should be studied independently. Methods of determining the haplotype of an individual are well known in the art and include, e.g., asymmetric PCR amplification (Newton et al. (1989) Nucleic Acids Res. 17:2503-2516; Wu et al. (1989) Proc.Natl. Acad. Sci. USA. 86:2757-2760), isolation of single chromosome by limited dilution followed by PCR amplification (Ruano et al. (1990) Proc. Natl.

Acad. Sci. USA. 87:6296-6300) and, for sufficiently close biallelic markers, double PCR amplification of specific alleles (Sarkar and Sommer, (1991) Biotechniques. 10:436-440).

Thus the present invention is further directed to the use of at least one UBP8rp-related biallelic marker of the present invention for determining the haplotype of an individual. For example, a method for determining a haplotype for a set of biallelic markers in an individual may comprise the steps of: a) genotyping said individual for at least one UBP8rp-related biallelic marker, b) genotyping said individual for a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker. In one embodiment, both markers are UBP8rp - related biallelic markers of the present ivention. In another embodiment, one marker is a UBP8rp related marker of the present invention and the other biallelic marker is biallelic marker 20 or 36.

Methods of determining a haplotype for a combination of more than two biallelic markers comprising at least one UBP8RP-related biallelic marker of the present invention in an individual are also encompassed by the present invention. In such methods, step (b) is repeated for each of the additional markers of the combination. Such a combination may comprise, e.g., 3, 4 or 5 biallelic markers.

When estimating haplotype frequencies in a population, one may use methods without assigning haplotypes to each individual. Such methods use a statistical method of haplotype determination. Thus another aspect of the present invention encompasses methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising the steps of: a) genotyping each individual in said population for at least one UBP8RP-related biallelic marker, b) genotyping each individual in said population for a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency. Such a method may also be performed for a combination of more than 2 biallelic markers. Step (c) may be performed using any method known in the art to determine or to estimate the frequency of a haplotype in a population. Preferably, a method based on an expectation-maximization (EM) algorithm (Dempster et al. (1977) JRSSB, 39:1-38; Excoffier and Slatkin, (1995) Mol Biol Evol. 12:921-7) leading to maximum-likelihood estimates of haplotype frequencies under the assumption of Hardy-Weinberg proportions (random mating) is used for performing step (c).

Having now fully described this invention, it will be appreciated by those skilled in the art that the same can be performed within a wide range of equivalent parameters without departing from the spirit and scope of the invention and without undue experimentation.

While this invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications. This application is intended to cover

any variations, uses or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth as follows in the scope of the appended claims.

All references cited herein, including journal articles or abstracts, published or unpublished patent application, issued patents or any other references, are entirely incorporated by reference herein, including all data, tables, figures and text presented in the cited references. Additionally, the entire contents of the references cited within the references cited herein are also entirely incorporated by reference.

Reference to known method steps, conventional methods steps, known methods or conventional methods is not any way an admission that any aspect, description or embodiment of the present invention is disclosed, taught or suggested in the relevant art.

The foregoing description of the specific embodiments will so fully reveal the general nature of the invention that others can, by applying knowledge within the skill of the art (including the contents of the references cited herein), readily modify and/or adapt for various application such specific embodiments, without undue experimentation, without departing from the general concept of the present invention. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed embodiments, based on the teaching and guidance presented herein. It is to be understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance presented herein, in combination with the knowledge of one of ordinary skill in the art.

EXAMPLES

Example 1: Identification of the UBP8rp gene

1. *Isolation of the UBP8rp mRNA*

RT-PCRs were performed on polyA+ RNAs from Stratagene (Reference Nos 778021 and 778022). cDNA was synthesized with the help of Rt-for-PCR kit (Clontech) with oligo(dT) and random primers for each RNA sample. The cDNA quantity obtained for the reactions varied between 0,6 and 1,5 ug of cDNA per reaction. A PCR reaction was performed using 5 µl out of 100 µl of the RT-PCR samples, primers of SEQ ID Nos. 5 and 6, and the rTth enzyme. The cycling was as follows: 94°C 5min ; 94°C 20sec,67°C 3min - 32 cycles ; 72°C 10min.

The first PCR reaction was diluted five fold and 2% thereof was used for performing a nested PCR reaction with primers of SEQ ID Nos. 7 and 8. The cycling conditions wer e identical as above.

The resulting product was sequenced using primers of SEQ ID Nos. 9-35. The sequencing was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing re actions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis (ABI Prism DNA Sequencing Analysis software (2.1.2 version)).

The cDNA comprised the Open Reading Frame of SEQ ID NO: 2. This Open Reading Frame codes for a 482 amino acid long protein, the UBP8rp protein (SEQ ID NO: 3).

2. *Identification and annotation of the UBP8rp gene*

The genomic region encoding the UBP8rp protein was ide ntified using bioinformatic tools. The UBP8rp gene is shown as SEQ ID NO: 1. This gene is located within the 10 -kb major region for susceptibility for psoriasis that was identified by Veal et al. (2002). The UBP8rp gene comprises two introns located at nucleotide positions 1018 to 1046 of SEQ ID NO: 1 and 1676 to 1718 of SEQ ID NO: 1 (see Figure 1).

Thus it has unexpectedly been found that a novel expressed gene is located within the 10-kb major region for susceptibility for psoriasis. In prior art litterat ure, the gene encoding UBP8rp was annotated as a silent pseudogene element not comprising any Open Reading Frame.

3. *Analysis of the UBP8rp protein*

The UBP8rp protein shows significant homology to the UBP8 ubiquitin isopeptidase. When comparing UBP8rp to UBP8 using the BLAST version 2.0 program (Altschul et al. 1990 J Mol Biol. 215:403-410), UBP8rp is found to be 74% identical to UBP8 (Figure 2). More specifically, amino acids 57 to 466 of UBP8rp show homology to amino acids 78 to 492 of UBP8 (81% of identity).

Using the SignalP and Toppred programs (Nielsen et al. 1999 Protein Engineering 12:3-9, von Heijne. 1992 J. Mol. Biol. 225:487-494), UBP8rp was found to be an intracellular protein. UBP8rp was further analyzed using the HMMER 2.1.1 program (Eddy. 1996 Current Opinion in Structural Biology 6:361-365). As shown on Figure 3, UBP8rp displays a rhodanese-like Pfam domain at amino acid positions 164 to 433 (score: 32.7; e-value: 8.3e-06). The presence of rhodanese-like domains is a common feature to UBPs since UBP8 and the ubiquitin isopeptidase 7 from *Saccharomyces cerevisiae* also display a single rhodanese-like domain.

Thus UBP8rp is a novel member of the UBP family. UBP8rp seems to belong to the ubiquitin-proteasome pathway, and may play a role in the selective degradation of intracellular proteins.

Example 2: Analysis of UBP8rp expression by quantitative PCR

The expression levels of UBP8rp in adult skin, fetal skin, testis, brain, adipose tissue, small intestine and colon was determined using commercial total RNA (Clontech). In addition, the expression levels of UBP8rp in adult skin was also determined using skin biopsies from l'Hôpital Pasteur (Paris, France).

20 µL of commercial total RNA were treated by 4 units of RNase free DNAse I (Ambion). The cDNA was obtained using the "Advantage RT for PCR" kit (Clontech) following the instructions provided by the supplier. The Quantitative PCR was performed using the TaqMan Universal PCR Master mix NO AmpErase UNG (Applied Biosystems). The reaction was performed with 25 ng of cDNA, 300nM of each primer and 200nM of Taqman probe. The program applied was: 40 to 50 cycles at 95°C for 10 minutes; 95°C for secondes; 60°C for 1 minute.

The experiments were performed on a 7900HT Applied Biosystems machine. Each experiment was performed either with primers of SEQ ID Nos. 36-38 or with primers of SEQ ID Nos. 39-41, as detailed in table 1 below. The efficiency of the chosen primers were calculated as decribed in the User Bulletin Applied Biosystems (1997 - updated 10/2001) ABI PRISM 7700 Sequence Detection System. Relative Quantification of Gene Expression. User Bulletin #2.

Table 1

	Primer forward	Primer reverse	Taqman MGB probe	Amplicon length
1st set of primers	SEQ ID NO: 36	SEQ ID NO: 37	SEQ ID NO: 38	98
2nd set of primers	SEQ ID NO: 39	SEQ ID NO: 40	SEQ ID NO: 41	109

The expression was calculated as described by Livak & Schmittgen (2001, Methods 25:402-408) The Ct is an absolute value indicating the relative expression level of a gene. A Ct under 20 is indicative of a highly expressed gene. A Ct between 35 and 40 is indicative of a weakly expressed gene. Calculation of the $2^{-\Delta\Delta Ct}$ value allows to compare expression levels of a gene in a target tissue to be studied and in a reference tissue.

In order to confirm that the primers specifically amplify UBP8rp and not another gene, the amplicons obtained by quantitative PCR were sequenced with the forward and reverse primers used for performing the QPCR. It was found that the cDNA amplified by PCR effectively corresponds to the UBP8rp cDNA.

The results of the quantitative expression analysis are shown in table 2. The $2^{-\Delta\Delta Ct}$ value was calculated using testis as a reference tissue, numerous genes being expressed in testis at high levels.

Table 2

	Primers					
	SEQ ID Nos. 36-38			SEQ ID Nos. 39-41		
Tissue	**efficiency**	**Ct**	$2^{-\Delta\Delta Ct}$	**efficiency**	**Ct**	$2^{-\Delta\Delta Ct}$
Testis	-	36.1	1.0	-	35.4	1.0
Adult skin (biopsy)	-	36.1	2.5	-	36.6	1.1
Adult skin (commercial)	-	39.6	-5.3	-	39.5	-7.8
Fetal skin	104%	36.7	-1.0	77%	37.9	-3.6
Brain	259%	38.9	-8.3	342%	38.7	-11.5
Adipose	-	38.3	-4.7	-	38.3	-7.0
Small intestin	-	38.3	-2.4	-	39.0	-6.1
Colon	-	37.2	-1.7	-	36.6	-1.8

Using commercially available RNA, UBP8rp is found to be significantly expressed in testis, foetal skin, and colon, allthough at a low level. In addition, UBP8rp is found to be expressed at a very low level in adult skin, brain, adipose and small intestin.

When using RNA from skin biopsies, UBP8rp is found to be significantly expressed in adult skin. Specifically, expression of UBP8rp is found to be higher in adult skin than in any other tissue, both with primers of SEQ ID Nos. 36-38 and with primers of SEQ ID Nos. 39-41.

When the experiment is performed using primers SEQ ID Nos. 36-38, expression of UBP8rp is found to be 2.5-fold higher in adult skin than in testis.

Example 3: <u>Identification of Biallelic markers located in the UBP8rp gene</u>

Eighteen biallelic markers were identified as detailed below (BM Nos. 17-19, 22, 25, 27-29, 31-35 and 37-41). Fifteen biallelic markers were identified using sequence data provided by Celera (BM Nos. 1, 2, 4, 6, 7, 10, 12-16, 21, 23, 24 and 26).

50 to 100 ng of genomic DNA from lymphoblastoid cell lines Lucy or Boleth (CEPH collection) were used to perform a PCR reaction with primers of SEQ ID Nos. 42 and 43. The PCR assays were performed using the following protocol:

- 5 units of Amplitaq enzyme (Perkin-Elmer, N°808-0101)
- 30 µl of reaction mix with 10X supplied Taq buffer
- 250 µM each dNTP,
- 15 µM of each primer
- cycling: 94°C 10 min, then 30 cycles of 3 steps – 94°C 30sec ; 55°C 30sec ; 72°C 30sec, then 72°C 10min

The PCR product was sequenced with the help of amplification primers. The sequences were blasted against genomic sequence, and sequence curves were compared. Biallelic marker Nos. 17-19 and 22 were thus identified.

50-100 ng of genomic DNA from lymphoblastoid cell lines Lucy or Boleth (CEPH collection) were used to perform a long-range PCR reaction with primers of SEQ ID Nos. 44 and 45. The PCR assays were performed using the following protocol:

- 2 units of rTTh XL enzyme (Perkin-Elmer)
- 50 µl of reaction mix with supplied 3,3X buffer and 200 µM of each dNTP
- 20 µM of each primer
- 1.1mM Mg0Ac.
- Cycling: 94°C 5min, then 32 cycles of 2 steps - 94°C 20sec ; 66°C 4min, then 72°C 10min.

The resulting product was sequenced with the help of the following pairs of primers: SEQ ID Nos. 46 and 47, SEQ ID Nos. 48 and 49, and SEQ ID Nos. 50 and 51. The sequences were compared by blast and by manual inspection of sequence electrophoregrams.

Sequencing by SEQ ID Nos. 46 and 47 allowed the identification of biallelic marker NO: 25. Sequencing by SEQ ID Nos. 48 and 49 allowed the identification of biallelic markers Nos.

27-29 and 31. Sequencing by SEQ ID Nos. 50 and 51 allowed the identification of biallelic markers Nos. 27-29 and 31.

The alternative alleles of biallelic markers Nos. 1-41 and their location within the UBP8rp gene are indicated in table 3. Biallelic markers Nos. 2-11 and 13-14 are located within the CDS of UBP8rp. Biallelic markers Nos. 2-8, 10 and 13-14 are coding SNPs. Biallelic markers Nos. 20 and 36, which are known to be highly associated with psoriasis, are shown in bold letters.

Table 3

BM No.	Internal Designation	Position on SEQ ID NO: 1	Exon No.	coding	Alternative Alleles	Sequence in:	
						Boleth	Lucy
1	hCV15819424	1199	-	-	A/G	-	-
2	hCV16030280	1262	2	Yes	C/T	-	-
3	SNP n.14	1426	2	Yes	C/G	-	-
4	hCV11691030	1444	2	Yes	G/T	-	-
5	SNP n.13	1487	2	Yes	A/G	-	-
6	hCV16030281	1490	2	Yes	A/G	-	-
7	hCV15819434	1505	2	Yes	G/T	-	-
8	SNP n.12	1518	2	Yes	C/T	-	-
9	SNP n.11	1554	2	No	C/T	-	-
10	hCV15819435	1630	2	Yes	A/G	-	-
11	SNP n.10	1638	2	No	A/T	-	-
12	hCV16030289	1680	- (splice site)	-	A/G	-	-
13	hCV16030290	1895	3	Yes	A/G	-	-
14	hCV16030297	2180	3	Yes	A/G	-	-
15	hCV16030298	2449	-	-	A/T	-	-
16	hCV16030299	2721	-	-	G/T	-	-
17	SNPG3127	3127	-	-	A/G	A/A	A/G
18	SNPG3137	3137	-	-	C/T	T/T	C/T
19	SNPG3138	3138	-	-	A/G	G/G	G/A
20	**SNP n.9**	**3183**	**-**	**-**	**A/G**	**A/A**	**A/G**
21	hCV15824895	3222	-	-	C/G	G/G	G/G
22	SNPG3269	3269	-	-	C/T	C/T	T/T
23	hCV15824896	3445	-	-	C/T	-	-
24	hCV16030306	3470	-	-	A/G	-	-
25	SNPG3915	3915	-	-	C/T	T/T	C/C
26	hCV16030307	3973	-	-	A/C	A/A	A/A
27	SNPG4254	4254	-	-	A/G	A/A	A/A
28	SNPG4472	4472	-	-	A/T	A/A	A/A
29	SNPG4660	4660	-	-	C/T	C/C	C/T
30	SNP n.8	4770	-	-	A/G	G/G	A/A
31	SNPG4919	4919	-	-	A/G	G/G	A/G
32	SNPG4973	4973	-	-	C/T	T/T	C/T
33	SNPG5063	5063	-	-	C/T	T/T	C/C
34	SNPG5065	5065	-	-	G/T	G/G	T/T

Table 3 (continued)

BM No.	Internal Designation	Position on SEQ ID NO: 1	Exon No.	coding	Alternative Alleles	Sequence in:	BM No.
35	SNPG5079	5079	-	-	C/T	C/C	C/T
36	**SNP n.7**	**5080**	-	-	**C/T**	**C/C**	**C/T**
37	SNPG5088	5088	-	-	C/G	G/G	C/C
38	SNPG5090	5090	-	-	C/T	T/T	C/C
39	SNPG5407	5407	-	-	C/T	C/C	C/T
40	SNPG5466	5466	-	-	A/G	G/G	G/G
41	SNPG5520	5520	-	-	C/T	T/T	T/T

REFERENCES

1. Altschul et al. (1990) J Mol Biol. 215:403-410
2. Borresen et al. (1988) Mutat Res. 202:77-83
3. Cather et al. (2003) Expert Opin Biol Ther. 3:361-370
4. Dempster et al. (1977) JRSSB, 39:1-38
5. Devereux J et al. (1984) Nucleic Acids Res. 12:387-395
6. Eddy. (1996) Current Opinion in Structural Biology 6:361-365
7. Elbashir et al. (2001) Nature 411:494-498
8. Ellington and Szostak (1990) Nature 346:818-822
9. Excoffier and Slatkin, (1995) Mol Biol Evol. 12:921-927
10. Farber et al. (1974) Arch Dermatol. 109:207-111
11. Fields and Song (1989) Nature. 340:245-6
12. Fredriksson et al. (1978) Dermatologica. 157:238-244
13. Fromont-Racine et al. (1997) Nat Genet. 16:277-282
14. Gnesutta et al. (2001) J Biol Chem. 276:39448-39454
15. Gottlieb et al. (2003) J Drugs Dermatol. 2:260-266
16. Grantham (1974) Science 185:862-864
17. Grompe et al. (1989) Proc Natl Acad Sci USA. 86:5888-5892
18. Kato et al. (2000) J Biol Chem. 275:37481-37487
19. Kim et al. (2003) Anal Biochem. 316:251-258
20. Lee et al. (2000) Am J Hum Genet. 67:1020-1024
21. Lessa et al. (1993) Mol Ecol. 2:119-129
22. Livak and Schmittgen (2001) Methods 25:402-408
23. Nair et al. (1997) Hum Mol Genet. 6:1349-1356
24. Naviglio et al. (1998) EMBO J. 17:3241-3250
25. Newton et al. (1989) Nucleic Acids Res. 17:2503-2516
26. Nielsen et al. (1999) Protein Engineering 12:3-9
27. Oka et al. (1999) Hum Mol Genet. 8:2165-2170
28. Oldenburg et al. (1992) Proc Natl Acad Sci U S A. 89:5393-5397
29. Orita et al. (1989) Proc Natl Acad Sci USA 86:2766-2770
30. Parmley and Smith (1988) Gene. 73:305-318
31. Pearson (1990) Methods in Enzymology, 183:63-99
32. Pearson and Lipman (1988) Proc Nat Acad Sci USA, 85:2444-2448
33. Ramunsen et al. (1997), Electrophoresis, 18: 588-598
34. Ruano et al. (1990) Proc. Natl. Acad. Sci. USA. 87:6296-6300
35. Sarkar and Sommer, (1991) Biotechniques. 10:436-440

36. Smith and Waterman (1981) J Mol Evol. 18:38-46
37. Trembath et al. (1997) Hum Mol Genet. 6:813-820
38. Valadon et al. (1996) J Immunol Methods. 197:171-179
39. Veal et al. (2001) J Med Genet. 38:7-13.
40. Veal et al. (2002) Am J Hum Genet. 71:554-564
41. von Heijne. (1992) J. Mol. Biol. 225:487-494
42. Wen et al. (2003) World J Gastroenterol. 9:1342-1346
43. Wilkinson (1997) FASEB J. 11:1245-1256
44. Wilkinson et al. (1995) Biochemistry 34:14535-14546
45. Wu et al. (1989) Proc.Natl. Acad. Sci. USA. 86:2757-2760
46. Zollner et al. (2002), J Clin Invest. 109:671-679

CLAIMS:

1. An isolated gene comprising introns having a sequence of:
    a) nucleotides 1018 to 1046 of SEQ ID NO: 1; and
    b) nucleotides 1676 to 1718 of SEQ ID NO: 1.
2. An isolated polynucleotide complementary to a messenger RNA transcribed from the gene of claim 1.
3. The polynucleotide of claim 2, wherein said polynucleotide comprises SEQ ID NO: 2 or a polynucleotide complementary thereto.
4. An isolated polypeptide encoded by the polynucleotide of any of claims 1 to 3.
5. The polypeptide of claim 4, wherein said polypeptide is selected from the group consisting of:
    a) a polypeptide comprising SEQ ID NO:3;
    b) a polypeptide comprising a span of at least 470 amino acids of SEQ ID NO: 3;
    c) a polypeptide comprising a span of at least 15 amino acids of SEQ ID NO: 3, wherein said span falls within amino acids 467 to 482 of SEQ ID NO: 3;
    d) a mutein of any of (a) to (c), wherein the amino acid sequence has at least 95%, 96%, 97%, 98% or 99% identity to at least one of the sequences in (a) to (c);
    e) a mutein of any of (a) to (c) which is encoded by a nucleic acid which hybridizes to the complement of a DNA sequence encoding any of (a) to (c) under highly stringent conditions; and
    f) a mutein of any of (a) to (c) wherein any changes in the amino acid sequence are conservative amino acid substitutions of the amino acid sequences in (a) to (c).
6. An expression vector comprising the gene of claim 1.
7. An expression vector comprising the polynucleotide of claim 2 or 3.
8. The expression vector of claim 6 or 7, wherein said polynucleotide encodes the polypeptide of claim 5.
9. The expression vector of any of claim 6 to 8, wherein said vector is a gene therapy vector.
10. A host cell comprising the expression vector of any of claims 6 to 9.
11. A method of making a polypeptide, said method comprising the steps of culturing a hos t cell according to claim 10 under conditions suitable for the production of a polypeptide of claim 4 or 5 within said host cell.

12. The method of claim 11, further comprising the step of purifying said polypeptide from the culture.

13. An antibody that specifically binds to the polypeptide of claim 4 or 5.

14. Use of a polypeptide of claim 4 or 5 as a target for screening for natural binding partners.

15. Use of the polypeptide of claim 4 or 5 as a target for screening candidate modulators.

16. The use of claim 15, wherein said candidate modulator is selected from the group consisting of a natural ligand, a small molecule, an aptamer, an antisense mRNA, a small interfering RNA and an antibody.

17. The use of claim 15 or 16, wherein said modulator is a candidate drug for the treatment of a chronic inflammatory disease.

18. The use of any of claims 14 to 17, wherein the activity of said polypeptide of claim 4 or 5 is assessed by measuring the de-ubiquitinating activity of said polypeptide of claim 4 or 5.

19. The use of any of claims 14 to 17, wherein the activity of said polypeptide of claim 4 or 5 is assessed by measuring the de-ubiquitinating activity of UBP8 in the presence of said polypeptide of claim 4 or 5.

20. Use of a modulator of a polypeptide of claim 4 or 5 for preparing a medicament for the treatment of a chronic inflammatory disease.

21. The use of claim 20, wherein said modulator is used in combination with a known drug for said chronic inflammatory disease.

22. The use of any of claims 17 to 21, wherein said chronic inflammatory disease is psoriasis.

23. A method of assessing the efficiency of a modulator of a polypeptide of claim 4 or 5 for the treatment of psoriasis, said method comprising administering said modulator to an animal model for psoriasis; wherein a determination that said modulator ameliorates a representative characteristic of psoriasis in said animal model indicates that said modulator is a drug for the treatment of psoriasis.

24. The method of claim 23, wherein said representative characteristic is a Psoriasis Area and Severity Index score.

25. The method of claim 24, wherein a 75% or greater improvement in Psoriasis Area and Severity Index scores (PASI 75) indicates that said modulator is a drug for the treatment of psoriasis.

26. The method of any of claims 23 to 25, wherein said animal model is a SCID-hu Mouse.

27. Use of at least one UBP8rp-related biallelic marker selected from the group consisting of the biallelic markers shown below for determining whether there is a significant association between said biallelic marker and a chronic inflammatory disease:

Biallelic maker No.	Position on SEQ ID NO: 1	Alternative nucleotides
1	1199	A/G
2	1262	C/T
4	1444	G/T
6	1490	A/G
7	1505	G/T
10	1630	A/G
12	1680	A/G
13	1895	A/G
14	2180	A/G
15	2449	A/T
16	2721	G/T
17	3127	A/G
18	3137	C/T
19	3138	A/G
21	3222	C/G
22	3269	C/T
23	3445	C/T
24	3470	A/G
25	3915	C/T
26	3973	A/C
27	4254	A/G
28	4472	A/T
29	4660	C/T
31	4919	A/G
32	4973	C/T
33	5063	C/T
34	5065	G/T
35	5079	C/T
37	5088	C/G
38	5090	C/T
39	5407	C/T
40	5466	A/G
41	5520	C/T

28. Use of at least one UBP8rp-related biallelic marker of the table set forth in claim 24 for diagnosing whether an individual suffers from or is at risk of suffering from a chronic inflammatory disease.

29. The use of claim 27 or 28, wherein said chronic inflammatory disease is psoriasis.

30. A method of genotyping comprising the steps of:

a) isolating a nucleic acid from a biological sample; and

b) detecting the nucleotide present at one or more of the UBP8rp-related biallelic markers of the table set forth in claim 27.

31. The method of claim 30, wherein said biological sample is derived from a single individual.

32. The method of claim 31, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said in dividual's genome.

33. The method of any of claims 30 to 32, wherein said determining is performed by a microsequencing assay.

34. The method of any of claims 30 to 33, further comprising amplifying a portion of a sequence comprising the biallelic marker prior to said determining step.

35. The method of claim 34, wherein said amplifying is performed by PCR.

36. The method of any of claims 31 to 35, further comprising the step of correlating the result of the genotyping steps with a risk of suffering from a chronic inflamma tory disease.

37. The method of claim 36, wherein said chronic inflammatory disease is psoriasis.

ABSTRACT

The invention encompasses a novel gene encoding a protein of the ubiquitin-proteasome pathway, UBP8rp. The invention also relates the use of UBP8rp polypeptides for screening for modulators, and to the use of said modulators for treating chronic inflammatory diseases such as, e.g., psoriasis, psoriatic arthritis, rheumatoid arthritis, asthma, inflammatory bowel disease and multiple sclerosis. The invention further relates to the use of biallelic markers located in the UBP8rp gene for diagnosing said chronic inflammatory diseases.

FIGURE 1A

```
Frame3   L  C  S  L  L  F  G  L  S  L  S  I  T  M  F  E  M  M  P
Frame2  V  M  *  S  F  V  W  S  L  P  *  H  N  D  V  *  D  D  A  I
Frame1  C  Y  V  V  F  C  L  V  S  P  L  A  *  R  C  L  R  *  C  H
DNA     TGTTATGTAGTCTTTTGTTTGGTCTCTCC CTTAGCATAACGATGTTTGAGATGATGCCAT 60

Frame3  F  I  H  F  C  C  *  A  A  A  E  Y  C  W  N  P  S  L  F  I
Frame2   H  S  F  L  L  L  S  S  C  R  V  L  L  E  S  Q  F  I  H  W
Frame1  S  F  I  F  V  A  E  Q  L  P  S  I  V  G  I  P  V  Y  S  L
DNA     TCATTCATTTTTGTTGCTGAGCAGCTGCCGAGTATTGTTGGAATCCCAGTTTATTCATTG 120

Frame3  G  F  C  V  S  S  *  *  T  C  G  F  L  Q  L  G  F  V  I  N
Frame2   F  L  C  L  Q  L  I  D  M  W  I  P  P  V  R  V  C  Y  *  *
Frame1  V  S  V  S  P  V  D  R  H  V  D  S  S  S  *  G  L  L  L  M
DNA     GTTTCTGTGTCTCCAGTTGATAGACATGTGGATTCCTCCAGTTAGGGTTTGTTATTAATG 180

Frame3  E  A  T  I  N  N  C  L  Q  V  W  T  Y  I  F  I  S  F  G  *
Frame2   S  H  Y  K  *  L  L  T  S  V  D  L  H  F  Y  F  F  W  I  N
Frame1  K  P  L  *  I  T  A  Y  K  C  G  L  T  F  L  F  L  L  D  K
DNA     AAGCCAC TATAAATAACTGCTTACAAGTGTGGACTTACATTTTTATTTCTTTTGGATAAA 240

Frame3  I  R  I  C  G  I  A  G  P  C  G  N  R  W  V  T  V  *  E  T
Frame2   T  Y  L  W  N  C  W  A  M  W  *  *  M  G  N  C  I  R  N  C
Frame1  Y  V  F  V  E  L  L  G  H  V  V  I  D  G  *  L  Y  K  K  L
DNA     TACGTATTTGTGGAATTGCTGGGCCATGTGGTAATAGATGGGTAACTGTATAAGAAACTG 300

Frame3  A  I  P  L  Y  K  L  A  A  T  F  F  A  F  L  P  A  I  S  D
Frame2   H  T  T  L  Q  I  G  C  H  I  F  C  I  P  T  S  N  I  R  H
Frame1  P  Y  H  F  T  N  W  L  P  H  F  L  H  S  Y  Q  Q  Y  Q  T
DNA     CCATACCACTTTACAAATTGGCTGCCACATTTTTTGCATTCCTACCAGCAATATCAGACA 360

Frame3  I  P  I  F  S  I  F  L  P  V  L  R  L  I  I  C  L  F  N  F
Frame2   S  Y  F  F  H  I  L  A  S  V  K  T  Y  H  M  S  F  *  L  Y
Frame1  F  L  F  F  P  Y  S  C  Q  C  *  D  L  S  Y  V  F  L  T  L
DNA     TTCCTATTTTTTCCATATTCTTGCCAGTGTTAAGACTTATCATATGTCTTTTTAACTTTA 420

Frame3  I  C  S  R  *  C  V  M  V  S  H  C  G  F  N  L  H  F  F  D
Frame2   L  L  *  V  M  C  D  G  F  S  L  W  F  *  L  A  L  L  *  *
Frame1  S  A  L  G  D  V  *  W  F  L  I  V  V  L  T  C  T  S  L  M
DNA     TCTGCTCTAGGTGATGTGTGATGGTTTCTCATTGTGGTTTTAACTTGCACTTCTTTGATG 480

Frame3  D  *  Y  C  L  L  S  F  H  V  H  L  S  D  L  L  H  I  F  Y
Frame2   L  V  L  F  A  I  F  S  C  S  S  K  R  L  I  T  Y  I  L  *
Frame1  T  S  I  V  C  Y  L  F  M  F  I  *  A  T  Y  Y  I  Y  F  M
DNA     ACTAGTATTGTTTGCTATCTTTTCATGTTCATCTAAGCGACTTATTACATATAT TTTATG 540

Frame3  E  L  F  C  K  F  N  D  *  F  Q  R  L  F  Q  N  S  L  V  F
Frame2   T  I  L  Q  I  Q  *  L  I  P  E  T  F  S  E  F  P  S  V  F
Frame1  N  Y  F  A  N  S  M  I  N  S  R  D  F  F  R  I  P  *  C  F
DNA     AACTATTTTGCAAATTCAATGATTAATTCCAGAGACTTTTTCAGAATTCCCTAGTGTTTT 600

Frame3  S  T  Y  T  M  K  L  V  T  K  K  D  F  H  F  F  L  S  Y  P
Frame2   Y  I  Y  N  E  V  G  D  K  E  R  L  S  F  L  P  F  L  S  I
Frame1  L  H  I  Q  *  S  W  *  Q  R  K  T  F  I  S  S  F  L  I  H
DNA     CTACATATACAATGAAGTTGGTGACAAAGAAAGACTTTCATTTCTTCCTTTCTTATCCAT 660

Frame3  L  I  F  F  L  L  K  L  L  L  F  G  R  D  E  V  S  L  I  R
Frame2   D  L  F  S  F  K  I  I  I  I  W  *  R  *  G  L  T  Y  Q  A
Frame1  *  S  F  F  F  *  N  Y  Y  Y  L  V  E  M  R  S  H  L  S  G
DNA     TGATCTTTTTTCTTTTAAAATTATTATTATT TGGTAGAGATGAGGTCTCACTTATCAGGC 720

Frame3  L  V  S  N  S  *  S  Q  V  I  L  P  P  Q  P  P  K  M  Q  G
Frame2   G  L  K  L  L  I  S  S  D  P  P  T  S  A  S  Q  N  A  G  I
Frame1  W  S  Q  T  P  D  L  K  *  S  S  H  L  S  L  P  K  C  R  D
DNA     TGGTCTCAAACTCCTGATCTCAAGTGATCCTCCCACCTCAGCCTCCCAAAATGCAGGGAT 780
```

FIGURE 1B

```
Frame3  L  Q  A  *  A  T  M  P  G  P  C  C  T  G  *  D  D  C  *  V
Frame2   T  G  M  S  H  H  A  W  S  L  L  H  W  L  G  *  L  L  G  V
Frame1  Y  R  H  E  P  P  C  L  V  L  V  A  L  V  R  M  T  V  R  C
DNA     TACAGGCATGAGCCACCATGCCTGGTCCTTGTTGCACTGGTTAGGATGACTGTTAGGTGT 840
.............................................................
Frame3  F  K  Q  E  *  *  E  L  T  C  L  F  T  R  N  L  N  K  F  T
Frame2   *  T  R  M  M  R  A  H  M  F  V  Y  K  E  L  K  Q  I  Y  K
Frame1  L  N  K  N  D  E  S  S  H  V  C  L  Q  G  T  *  T  N  L  Q
DNA     TTAAACAAGAATGATGAGAGCTCACATGTTTGTTTACAAGGAACTTAAACAAATTTACAA 900
                  |>START Orf
.............................................................
Frame3  R  K  K  P  I  P  I  K  K  W  A  K  D  I  N  R  H  F  S  E
Frame2   K  K  T  H  P  H  Q  K  V  G  K  G  Y  K  Q  T  L  L  R  G
Frame1  E  K  N  P  S  P  S  K  S  G  Q  R  I  *  T  D  T  S  Q  R
DNA     GAAAAAAACCCATCCCCATCAAAAAGTGGGCAAAGGATATAAACAGACACTTCTCAGAGG 960
.............................................................
Frame3  E  D  I  Y  V  A  K  K  H  M  K  K  S  S  H  T  Y  M  K  R
Frame2   R  H  L  R  G  Q  E  T  Y  E  K  K  L  T  H  V  Y  E  T  *
Frame1  K  T  F  T  W  P  R  N  I  *  K  K  A  H  T  R  I  *  N  V
                                  begin of homology orf to UBP8
DNA     AAGACATTTACGTGGCCAAGAAACATATGAAAAAAAGCTCACACACGTATATGAAACGTG 1020
                                               End 1st exon ||>intron1
.............................................................
Frame3  D  C  L  *  S  Y  P  K  K  N  *  F  Q  A  T  A  V  L  L  P
Frame2   L  F  I  I  L  S  K  K  E  L  I  S  S  N  S  S  I  T  S  I
Frame1  T  V  Y  N  P  I  Q  K  R  T  D  F  K  Q  Q  Q  Y  Y  F  H
DNA     ACTGTTTATAATCCTATCCAAAAAAGAACTGATTTCAAGCAACAGCAG TATTACTTCCAT 1080
                                 |> begin 2nd exon Orf
                                  start L1 repeat
.............................................................
Frame3  F  N  T  W  T  C  K  H  Q  K  S  H  W  R  N  *  T  T  L  *
Frame2   Q  Y  L  D  L  Q  T  S  K  K  P  L  E  K  L  N  D  S  L  K
Frame1  S  I  L  G  P  A  N  I  K  K  A  T  G  E  T  E  R  L  S  E
DNA     TCAATACTTGGACCTGCAAACATCAAAAAAGCCACTGGAGAAACTGAACGACTCTCTGAA 1140
.............................................................
Frame3  K  P  *  T  K  I  *  R  S  *  N  L  E  K  T  *  G  K  G  Q
Frame2   A  L  N  *  D  M  K  K  L  K  S  G  K  N  L  R  K  R  T  G
Frame1  S  L  K  L  R  Y  E  E  V  E  I  W  K  K  L  E  E  K  D  R
DNA     AGCCTTAAA CTAAGATATGAAGAAGTTGAAATCTGGAAAAAACTTGAGGAAAAGGACAGG 1200
                  ! end of L1 repeat                              R
hCV15819424
.............................................................
Frame3  A  G  G  S  T  V  A  T  T  K  K  A  G  N  R  K  R  G  W  Q
Frame2   R  G  K  H  S  G  Y  N  K  K  G  R  K  Q  E  E  R  M  A  A
Frame1  Q  G  E  A  Q  W  L  Q  Q  K  R  Q  E  T  G  R  E  D  G  S
DNA     CAGGGGGAAGCACAGTGGCTACAACAAAAAAGGCAGGAAACAGGAAGAGAGGATGGCAGC 1260
.............................................................
Frame3  H  V  G  *  R  F  F  G  D  C  I  G  F  Q  R  Q  N  P  K  E
Frame2   C  W  L  K  V  L  W  R  L  Y  W  I  P  K  T  K  P  K  R  A
Frame1  M  L  A  K  G  S  L  E  I  V  L  D  S  K  D  K  T  Q  K  S
DNA     ATGTTGGCTAAAGGTTCTTTGGAGATTGTATTGGATTCCAAAGACAAAACCCAAAAGAGC 1320
         Y hCV16030280 (M/T) confirmed Boleth
.............................................................
Frame3  Q  W  *  K  E  *  K  M  *  D  Q  R  E  R  S  N  H  S  K  G
Frame2   M  V  K  R  M  K  N  V  R  P  K  R  K  E  Q  S  Q  Q  R  N
Frame1  N  G  E  K  N  E  K  C  E  T  K  E  K  G  A  I  T  A  K  E
DNA     AATGGTGAAAAGAATGAAAAATGTGAGACCAAAGAGAAAGGAGCAATCACAGCAAAGGAA 1380
.............................................................
Frame3  T  I  H  N  D  D  G  *  K  H  Q  L  D  Y  N  G  C  S  K  N
Frame2   Y  T  Q  *  *  W  I  K  T  S  A  *  L  *  W  M  L  K  E  C
Frame1  L  Y  T  M  M  M  D  K  N  I  S  L  I  I  M  D  A  Q  R  M
DNA     CTATACACAATGATGA TGGATAAAAACATCAGCTTGATTATAATGGATGCTCAAAGAATG 1440
                      Begin of homology of orf1 to UBP8
                                                |
                                              C/G SNP14 C:allele1
.............................................................
```

FIGURE 1C

```
Frame3  A  G  L  S  G  F  L  Y  F  T  F  S  Q  C  S  *  K  S  H  Q
Frame2   R  I  I  R  I  P  V  F  Y  I  L  S  V  F  L  K  K  P  S  V
Frame1  Q  D  Y  Q  D  S  C  I  L  H  S  L  S  V  P  E  K  A  I  S
DNA     CAGGATTATCAGGATTCCTGT ATTTTACATTCTCTCAGTGTTCCTGAAAAAGCCATCAGT 1500
                                                     R    hCV16030281
          K hCV11691030
                                                    R SNP13 G:allele 1

Frame3  S  R  S  H  C  *  L  D  *  S  T  P  P  R  *  F  Y  R  Y  M
Frame2   Q  E  S  L  L  A  G  L  K  H  T  S  Q  M  I  L  *  I  H  G
Frame1  P  G  V  T  A  S  W  I  E  A  H  L  P  D  D  S  I  D  T  W
DNA     CCAGGAGTCACTGCTAGCTGGATTGAAGCACACCTCC CAGATGATTCTATAGATACATGG 1560
          KhCV15819434 Y hCV16030288(SNP12)  T:allele1      Y SNP11 C:
allele 2

Frame3  E  E  E  G  E  C  G  V  Y  G  T  S  *  L  V  *  F  C  K  R
Frame2   R  R  G  G  M  W  S  I  W  Y  F  L  T  G  L  V  L  Q  K  I
Frame1  K  K  R  G  N  V  E  Y  M  V  L  L  D  W  F  S  S  A  K  D
DNA     AAGAAGAGGGGGAATGTGGAGTATATGGTACTTCTTGACTGGTTTAGTTCTGCAAAAGAT 1620

Frame3  F  T  D  W  N  N  T  L  A  S  E  R  C  T  F  Q  V  G  K  *
Frame2   Y  R  L  E  Q  H  S  G  I  *  K  M  H  F  S  S  G  K  V  K
Frame1  L  Q  I  G  T  T  L  W  H  L  K  D  A  L  F  K  W  E  S  K
DNA     TTACAGATTGGAACAACACTCTGG CATCTGAAAGATGCACTTTTCAAGTGGGAAAGTAAA 1680
                                                 End exon 2 |
                     R hCV15819435               s456>               R
hCV16030289 splice site
                          W hCV15819436 SNP10 T:allele 1

Frame3  N  C  P  V  Q  W  A  L  G  L  W  F  *  R  E  A  I  K  T  G
Frame2   L  S  C  A  M  G  L  G  P  L  V  L  E  G  G  Y  K  N  W  F
Frame1  T  V  L  C  N  G  P  W  A  F  G  F  R  G  R  L  *  K  L  V
DNA     ACTGTCCTGTGCAATGGGCCTTGGGCCTTTGGTTTTAGAGGGAGGCTATAAAAACTGGTT 1740
                                                 |>begin exon3

Frame3  S  F  A  I  P  S  I  Q  Q  M  L  R  S  L  H  P  H  N  T  R
Frame2   L  C  Y  S  Q  Y  T  T  N  A  K  V  T  P  P  P  Q  H  Q  N
Frame1  P  L  L  F  P  V  Y  N  K  C  *  G  H  S  T  P  T  T  P  E
DNA     CCTTTGCTATTCCCAGTATACAACAAATGCTAAGGTCACTCCACCCCCACAACACCAGAA 1800
                    <R541

Frame3  M  K  S  C  L  S  H  W  I  L  L  I  P  H  W  K  N  Q  F  L
Frame2   E  E  L  S  I  S  L  D  F  T  Y  P  S  L  E  E  S  I  P  S
Frame1  *  R  V  V  Y  L  I  G  F  Y  L  S  L  I  G  R  I  N  S  F
DNA     TGAAGAGTTGTCTATCTCATTGGATTTTACTTATCCCTCATTGGAAGAATCAATTCCTTC 1860

Frame3  L  N  L  L  P  R  C  H  L  H  L  *  K  W  M  K  T  *  N  *
Frame2   K  P  A  A  E  M  P  P  P  P  I  K  V  D  E  D  I  E  L  I
Frame1  *  T  C  C  R  D  A  T  S  T  Y  K  S  G  *  R  H  R  I  D
DNA     TAAACCTGCTGCCGAGATGCCACCTCCACCTATAAAAGTGGATGAAGACATAGAATTGAT 1920
                                        R hCV16030290 K/E confirmed

Frame3  *  V  I  K  *  V  I  M  I  K  M  R  G  Q  D  H  *  I  Y  Q
Frame2   S  D  Q  I  S  D  N  D  Q  N  E  R  T  G  P  L  N  I  S  I
Frame1  K  *  S  N  K  *  *  *  S  K  *  E  D  R  T  T  E  Y  I  N
DNA     AAGTGATCAAATAAGTGATAATGATCAAAATGAGAGGACAGGACCACTGAATATATCAAT 1980

Frame3  F  Q  L  N  Q  L  L  L  L  N  L  M  F  H  P  S  F  S  Q  C
Frame2   P  V  E  S  V  A  A  S  K  S  D  V  S  P  I  I  Q  P  V  P
Frame1  S  S  *  I  S  C  C  F  *  I  *  C  F  T  H  H  S  A  S  A
DNA     TCCAGTTGAATCAGTTGCTGCTTCTAAATCTGATGTTTCACCCATCATTCAGCCAGTGCC 2040

Frame3  L  A  *  R  M  F  H  R  L  I  I  L  K  N  W  Q  S  N  C  L
Frame2   S  I  K  N  V  P  Q  I  D  H  T  K  K  L  A  V  K  L  P  E
Frame1  *  H  K  E  C  S  T  D  *  S  Y  *  K  T  G  S  Q  I  A  *
DNA     TAGCATAAAGAATGTTCCACAGATTGATCATACTAAAAAACTGGCAG TCAAATTGCCTGA 2100
```

FIGURE 1D

```
Frame3  K  S  I  *  S  N  L  K  V  Q  I  M  S  N  S  L  L  R  M  K
Frame2  E  H  I  I  K  S  E  S  T  N  H  E  Q  Q  S  P  Q  N  E  K
Frame1  R  A  Y  N  Q  I  *  K  Y  K  S  *  A  T  V  S  S  E  *  K
DNA     AGAGCATATAATCAAATCTGAAAGTACAAATCATGAGCAACAGTCTCCTCAGAATGAAAA 2160

Frame3  K  L  F  L  I  V  P  P  S  Q  *  F  P  L  Q  L  S  C  *  Q
Frame2  V  I  P  D  C  S  T  K  P  V  V  S  S  P  T  L  M  L  T  D
Frame1  S  Y  S  *  L  F  H  Q  A  S  S  F  L  S  N  S  H  V  N  R
DNA     AGTTATTCCTGATTGTTCCACCAAGCCAGTAGTTTCCTCTCCAACTCTCATGTTAACAGA 2220
                          R hCV160 30297 T/A E2F-1

Frame3  M  K  K  R  L  I  F  M  Q  K  L  L  F  *  W  R  K  T  N  K
Frame2  E  E  K  A  H  I  H  A  E  T  A  L  L  M  E  K  N  K  Q  E
Frame1  *  R  K  G  S  Y  S  C  R  N  C  S  S  N  G  E  K  Q  T  R
DNA     TGAAGAAAAGGCTCATATTCATGCAGAAACTGCTCTTCTAATGGAGAAAAACAAACAAGA 2280

Frame3  K  K  N  F  R  K  D  S  K  G  N  R  K  K  L  R  R  E  E  H
Frame2  K  E  L  Q  E  R  Q  Q  G  K  Q  K  E  T  E  E  G  R  T  R
Frame1  K  R  T  S  G  K  T  A  R  E  T  E  R  N  *  G  G  K  N  T
DNA     AAAAGAACTTCAGGAAAGACAGCAAGGGAAACAGAAAGAAACTGAGGAGGGAAGAACACG 2340

Frame3  E  Q  K  A  K  K  K  Q  E  A  E  E  N  E  I  T  Q  K  Q  Q
Frame2  A  K  S  Q  K  E  T  R  S  *  R  K  *  N  Y  T  E  A  T  K
Frame1  S  K  K  P  K  R  N  K  K  L  K  K  M  K  L  H  R  S  N  K
DNA     AGCAAAAAGCCAAAAAGAAACAAGA AGCTGAAGAAAATGAAATTACACAGAAGCAACAAA 2400
                                    | END of orf

Frame3  K  A  K  E  E  M  E  K  K  E  R  E  Q  A  K  K  E  D  K  E
Frame2  S  K  R  R  N  G  E  E  R  T  *  T  G  Q  E  R  G  *  R  N
Frame1  K  Q  K  K  K  W  R  R  K  N  V  N  R  P  R  K  R  I  K  K
DNA     AAGCAAAAGAAGAAATGGAGAAGAAAGAACGTGAACAGGCCAAGAAAGAGGATAAAGAAA 2460
                                                     W hCV16030298

Frame3  I  S  A  K  K  G  K  E  I  T  R  V  K  R  Q  S  K  S  D  H
Frame2  L  S  K  E  G  Q  R  N  N  K  S  K  K  T  K  *  K  *  S  *
Frame1  S  Q  Q  R  R  A  K  K  *  Q  E  *  K  D  K  V  K  V  I  M
DNA     TCTCAGCAAAGAAGGGCAAAGAAATAACAAGAGTAAAAAGACAAAGTAAAAGTGATCATG 2520

Frame3  E  T  S  G  A  E  K  S  V  E  D  R  G  R  R  C  S  T  P  E
Frame2  N  L  W  C  R  E  V  C  R  G  Q  G  E  K  M  F  N  P  R  S
Frame1  K  P  L  V  P  R  S  L  *  R  T  G  G  E  D  V  Q  P  Q  K
DNA     AAACCTCTGGTGCCGAGAAGTCTGTAGAGGACAGGGGGAGAAGATGTTCAACCCCAGAAG 2580

Frame3  V  Q  K  K  S  T  R  D  V  S  H  T  S  A  T  G  D  S  G  S
Frame2  T  E  K  V  N  K  R  C  V  P  Y  I  C  D  R  G  F  R  F  R
Frame1  Y  R  K  S  Q  Q  E  M  C  P  I  H  L  R  Q  G  I  Q  V  Q
DNA     TACAGAAAAAGTCAACAAGAGATGTGTCCCATACATCTG CGACAGGGGATTCAGGTTCAG 2640

Frame3  G  K  P  F  K  I  K  G  Q  P  E  T  G  I  L  R  T  E  T  F
Frame2  Q  A  F  *  D  *  R  T  T  R  N  W  N  S  K  D  R  N  F  *
Frame1  A  S  L  L  R  L  K  D  N  Q  K  L  E  F  *  G  Q  K  L  L
DNA     GCAAGCCTTTTAAGATTAAAGGACAACCAGAAACTGGAATTCTAAGGACAGAAACTTTTA 2700

Frame3  R  E  D  T  D  D  T  *  R  N  K  T  Q  R  E  P  S  I  I  A
Frame2  R  G  Y  R  *  Y  L  K  K  *  N  S  T  R  T  F  D  N  S  T
Frame1  E  R  I  Q  M  I  L  K  E  I  K  L  N  E  N  L  R  *  *  H
DNA     GAGAGGATACAGATGATACTTAAAGAAATAAAACTCAACGAGAACCTTCGATAATAGCAC 2760
                          K hCV16030299

Frame3  R  S  E  E  M  G  R  M  V  P  G  L  P  S  G  W  A  K  F  L
Frame2  K  *  R  N  G  E  D  G  T  R  T  A  F  R  L  G  Q  V  S  *
Frame1  E  V  K  K  W  G  G  W  Y  Q  D  C  L  Q  A  G  P  S  F  L
DNA     GAAGTGAAGAAATGGGGAGGATGGTACCAGGACTGCCTTCAGGCTGGGCCAAGTTTCTTG 2820
```

FIGURE 1E

```
Frame3  D P I T G T F H Y Y H S P L T L F I C T
Frame2  S N H W N V S L L S F T T N T V H M Y P
Frame1  I Q S L E R F I I I I H H * H C S Y V P
DNA     ATCCAATCACTGGAACGTTTCATTATTATCATTCACCACTAACACTGTTCATATGTACCC 2880

Frame3  H W K W L L H L H L L P P L Q L I K A S
Frame2  L E M A P S S A P P S T P P T H K G K P
Frame1  T G N G S F I C T S F H P S N S * R Q A
DNA     ACTGGAAATGGCTCCTTCATCTGCAC CTCCTTCCACCCCTCCAACTCATAAAGGCAAGCC 2940

Frame3  H R F L L S R I G N L P N * N A L T P P
Frame2  Q I P A K Q D R E P S K L K C S Y S S P
Frame1  T D S C * A G * G T F Q T E M L L L L P
DNA     ACAGATTCCTGCTAAGCAGGATAGGGAACCTTCCAAACTGAAATGCTCTTACTCCTCCCC 3000

Frame3  Q I * P R L F K R K P A V T P T V N Q E
Frame2  D I T Q A I Q E E A S S N S N S * S G R
Frame1  R Y N P G Y S R G S Q Q * L Q Q L I R K
DNA     AG ATATAACC CAGGCTATTCAAGAGGAAGCCAGCAGTAACTCCAACAGTTAATCAGGAAG 3060

Frame3  D K P T C Y P K A E I S R L S A S Q I W
Frame2  Q A N M L P * S * D L K A F C F S D L E
Frame1  T S Q H A T L K L R S Q G F L L L R F G
DNA     ACAAGCCAACATGCTACCCTAAAGCTGAGATCTCAAGGCTTTCTGCTTCTCAGATTTGGA 3120

Frame3  K L N P V F G G S G P A L T G L R N L G
Frame2  T Q S C F W R F W T S S Y W T S * L R K
Frame1  N S I L F L E V L D Q L L L D F V T * E
DNA     AACTCAATCCTGTTTTTGGAGGTTCTGGACCAGCTCTTACTGGACTTCGTAACTTAGGAA 3180
        | SNPG3127 A/G lucy, A/A Boleth
        | SNPG3137 C/T Lucy, T /T Boleth
        | SNPG3138 G/A Lucy G/G Boleth

Frame3  N T C Y M N S I L Q C L C N P P H L A D
Frame2  Y L L Y E L N I A V P M * S S T F G * L
Frame1  I L V I * T Q Y C S A Y V I L H I W L I
DNA     ATACTTGTTATATGAACTCAATATTGCAGTGCCTATGTAATCCTCCACATTTGGCTGATT 3240
        G SNP9: A allele 2 (associated) AG Lucy, AA Boleth
        S hC V15824895 GG Lucy, GG
boleth

Frame3  Y F N R N C Y Q D D I N K S N L L G A *
Frame2  F Q P K L L S G * Y * Q V K F V R G I K
Frame1  I S T E T V I R M I L T S Q I C * G H K
DNA     ATTTCAACCGAAACTGTTATCAGGATGATATTAACAAGTCAAATTTGTTAGGGGCATAAA 3300
        | SNPG3269 TC boleth , TT Lucy

Frame3  R * S G R R I W Y N H E S P V D R T V *
Frame2  V K W Q K N L V * S * K P R G Q D S I D
Frame1  G E V A E E F G I I M K A P W T G Q Y R
DNA     GGTGAAG TGGCAG AAGAATTTGGTATAATCATGAAAGCCCCGTGGACAGGACAGTATAGA 3360

Frame3  I Y Q S K R P * S H H W E D Q L P V C R
Frame2  I S V Q K T L K S P L G R S I T S L Q D
Frame1  Y I S P K D L K V T I G K I N Y Q F A G
DNA     TATATCAGTCCAAAAGACCTTAAAGTCACCATTGGGAAGATCAATTACCAGTTTGCAGGA 3420

Frame3  I Q S R F T R I S S V P N G W S P * R S
Frame2  T V K I H K N F F C S * W M V S M K I *
Frame1  Y S Q D S Q E F L L F L M D G L H E D L
DNA     TACAGTCAAGATTCACAAGAATTTCTTCTGTTCCTAATGGA TGGTCTCCATGAAGATCTA 3480
        Y hCV15824896        R hCV16030306
```

FIGURE 1F

```
Frame3 K * N * * S E D I * R R K * * S S Q * L
Frame2 I K L I I G R H I K K K I M I I S M T L
Frame1 N K T D N R K T Y K E E N N D H L N D F
DNA AATAAAACTGATAATCGGAAGACATATAAAGAAGAAAATAATGATCATCTCAATGACTTT 3540

Frame3 * S C R T C L A E T Q A A L * V Y Y C C
Frame2 K L Q N M P G R N T S G S M S L L L L H
Frame1 K A A E H A W Q K H K R L Y E S I I V A
DNA AAAGCTGCAGAACATGCCTGGCAGA AACACAAGCGGCTCTATGAGTCTATTATTGTTGCA 3600

Frame3 T F S G S I Q I Y S T V P H P S Q K V *
Frame2 F F R V N S N L Q Y S A S P V T K S L G
Frame1 L F Q G Q F K S T V Q C L T R H K K S R
DNA CTTTTTCAGGGTCAATTCAAATCTACAGTACAGTGCCTCACCCGTCACAAAAAGTCTAGG 3660

Frame3 D T * G L H V F V S T D S I H K * M Y I
Frame2 H L R P S C I C L Y * * H P Q V N V H Y
Frame1 T L E A F M Y L S L L I A S T S K C T L
DNA ACACTTGAGGCCTTCAT GTATTTGTCTCTACTGATAGCATCCACAA GTAAAT GTACATTA
3720

Frame3 I G L P * I I F * R R K T H R * * Q I L
Frame2 R I A L D Y F L K K K N S * I I T D F T
Frame1 * D C L R L F S K E E K L I D N N R F Y
DNA TAGGATTGCCTTAGATTATTTTCTAAAGAAGAAAAACTCATAG ATAATAACAG ATTTTAC 3780

Frame3 L Q S L Q S S T G F L K K K S G S Y H L
Frame2 A I F A E L D G I L K K E I W K L P P V
Frame1 C N L C R A R R D S * K R N L E V T T C
DNA TGCAATCTTTGCAGAGCTCGACGGGATTCTTAAAAAAGAAATCTGGAAGTTACCACCTGT 3840

Frame3 C F * C I * N I F P T M A G G N K N Y R
Frame2 L L V H L K H F S Y N G R W K Q K L Q T
Frame1 A F S A S E T F F L Q W Q V E T K I T D
DNA GCTTTTAGTGCATCTGAAACATTTTTCCTACAATGGCAG GTGGAAACAAAAATTACAGAC 3900

Frame3 H L W T S R * K I L P C H S M L L V Q R
Frame2 S V D F P L E N L A L S Q Y V I G P K N
Frame1 I C G L P V R K S C L V T V C Y W S K E
DNA ATCTGTGGACTTCCCGTTAGAAAATCTTGCCTTGTCACAGTATGTTATTGGTCCAAAGAA 3960
Y SNPG3915 CC Lucy, TT Boleth

Frame3 T I * R N I I C F L F Q I T A V G W M E
Frame2 N L K K Y N L F S V S D H C G G L D G G
Frame1 Q F E E I * F V F C F R S L R W A G W R
DNA CAATTTGAAGAAATATAATTTGTTTTCTGTTTCAGATCACTGCGGTGGGCTGGAT GGAGG 4020
M hCV16030307 AA Lucy, AA Boleth

Frame3 A I T Q P T V K M Q Q N S G G L S L M I
Frame2 H Y T A Y C K N A A K Q R W F K F D D H
Frame1 P L H S L L * K C S K T A V V * V * * S
DNA CCATTACACAGCCTACTGTAAAAATGCAGCAAAACAGCGGTGGTTTAAGTTTGATGATCA 4080

Frame3 M K F L I S L F L L * N L Q Q L I S S F
Frame2 E V S D I S V S S V K S S A A Y I L F Y
Frame1 * S F * Y L C F F C E I F S S L Y P L L
DNA TGAAGTTTCTGATATCTCTGTTTCTTCTGTGAAATCTTCAGCAGCTTATATCCTCTTTTA 4140

Frame3 I L L W D H E * L M * A H K E T * V I N
Frame2 T S L G P * V T D V G T * G D I G Y K L
Frame1 Y F F G T M S N * C R H I R R H R L * T
DNA TACTTCTTTGGGACCATGAGTAACTGATGTAGGCACATAAGGAGACATAGGTTATAAACT 4200
```

FIGURE 1G

```
Frame3  * L S F K R L S N T I L E M L I K I V V
Frame2   V I F * K A Q Q H N S * N A Y Q D S G
Frame1  S Y L L K G S A T Q F L K C L S R * W *
DNA     AGTTATCTTTTAAAAGGCTCAGCAACACAATTCTTGAAATGCTTATCAAGATAGTGGTAG 4260
                                                          | AA  lucy, AA
Boleth
.....................................................................
Frame3  A I A G H L E E F * D S G S C V T S T I
Frame2   N S W P F R G I L G Q W E L C Y * H Y I
Frame1  Q * L A I * R N S R T V G A V L L A L Y
DNA     CAATAGCTGGCCATTTAGAGGAATTCTAGGACAGTGGGAGCTGTGTTACTAGCACTATAT 4320
.....................................................................
Frame3  * F L S V V T N N T * Q V L Q * A S L T
Frame2   I P V S G D K * H L T S I A V S I T Y R
Frame1  N S C Q W * Q I T L N K Y C S K H H L Q
DNA     AATTCCTGTCAGTGGTGACAAATAACACTTAACAAGTATTGCAGTAAGCATCACTTACAG 4380
.....................................................................
Frame3  G T I Y F K T T F L V C S K V K I I N *
Frame2   Y H L F Q N N F F S L L Q S * N N * L A
Frame1  V P F I S K Q L F * S A P K L K * L T S
DNA     GTACCATTTATTTCAAAACAACTTTTTTAGTCTGCTCCAAAGTTAAAATAATTAACTAGC 4440
.....................................................................
Frame3  L S I I I L L V * K P L Y P F F P F H C
Frame2   K H Y Y S T G L K T F V P F F S F S L L
Frame1  * A L L F Y W S K N L C T L F F L F T V
DNA     TAAGCATTATTATTCTACTGGTCTAAAAACCTTTGTACCCTTTTTTTCCTTTTCACTGTT 4500
                                  | AA Lucy, AA Boleth
.....................................................................
Frame3  Y S L F T F L N P I F I Y Y E Y S R M M
Frame2   Q P F H I S K S H L H I L * I L * N D V
Frame1  T A F S H F * I P S S Y T M N T L E * C
DNA     ACAGCCTTTTCACATTTCTAAATCCCATCTTC ATATACTATGAATACTCTAGAATGATGT 4560
.....................................................................
Frame3  * S R * E C M C T Y L L H T Y T S N R Y
Frame2   K Q I G M Y V Y I F I A Y L H I K S I Y
Frame1  E A D R N V C V H I Y C I L T H Q I D I
DNA     GAAGCAGATAGGAATGTATGTGTACATATTTATTGCATACTTACACATCAAATCGATATA 4620
.....................................................................
Frame3  T * F N M W S F R E T * N S E D C I F F
Frame2   I V * H V V L S * N L E L R G L H F F L
Frame1  H S L T C G P F V K L R T Q R I A F F S
DNA     CATAGTTTAACATGTGGTCCTTTCGTGAAACTTAGAACTCAGAGGATTGCATTTTTTTCT 4680
                                           |SNPG4660 CC Boleth, CT Lucy
.....................................................................
Frame3  F E H I L S N C S A F L G K * Q G K A I
Frame2   * A Y F E * L Q C F L R E M T G Q S Y F
Frame1  L S I F * V T A V L S * G N D R A K L F
DNA     TTGAGCATATTTTGAGTAACTGCAGTGCTTTCTTAGGGAAATGACAGGGCAAAGCTATTT 4740
.....................................................................
Frame3  F L L A L G A F G C A K S L S * K I N G
Frame2   S V G F G G I W V R * I F I L K N K W K
Frame1  F C W L W G H L G A L N L Y L K K * M E
DNA     TTCTGTTGGCTTTGGGGGCATTTGGGTGCGCTAAATCTTTATCTTAAAAAATAAATGGAA 4800
                                  |G/A SNP8  G:all GG Boleth AA Lucy
.....................................................................
Frame3  N F L * F F K M R H * N L N E K N L K S
Frame2   L P L I F * N E T L K S * * E K F K K L
Frame1  T S F N F L K * D I K I L M R K I * K A
DNA     ACTTCCTTTAATTTTTTAAAATGAGACATTAAAATCTTAATGAGAAAAATTTAAAAAGCT 4860
                                    Homology to 3/ UBP8
.....................................................................
```

FIGURE 1H

```
Frame3  S  I  S  L  L  I  R  E  M  *  I  K  A  T  M  R  Y  H  L  P
Frame2   N  I  T  A  H  *  R  N  V  N  Q  S  H  N  E  I  P  S  P  T
Frame1    Q  Y  H  C  S  L  E  K  C  K  S  K  P  Q  *  D  T  I  S  H
DNA     CAATATCACTGCTCATTAGAGAAATGTAAATCAAAGCCACAATGAGATACCATCTCCCAC 4920
                                                               |AG Lucy
GG Boleth SNPG4919

Frame3  P  V  R  M  V  I  I  K  K  S  R  N  N  R  C  W  *  G  C  G
Frame2   S  Q  N  G  N  Y  *  K  V  K  K  Q  *  M  L  V  R  L  W  R
Frame1    Q  S  E  W  *  L  L  K  S  Q  E  T  I  D  A  G  E  A  V  E
DNA     CAGTCAGAATGGTAATTATTAAAAAGTCAAGAAACAATAGATGCTGGTGAGGCTGTGGAG 4980
                                                           | C/T Lucy TT
Boleth  SNPG4973

Frame3  E  I  G  T  L  L  H  C  C  W  E  C  K  L  V  Q  P  L  W  K
Frame2   N  R  N  T  F  T  L  L  L  G  M  *  T  S  S  T  I  V  E  D
Frame1    K  *  E  H  F  Y  T  V  V  G  N  V  N  *  F  N  H  C  G  R
DNA     AAATAGGAACACTTTTACACTGTTGTTGGGAATGTAAACTAGTTCAACCATTGTGGAAGA 5040

Frame3  T  V  W  P  F  L  R  D  L  E  P  E  I  L  F  D  P  L  G  I
Frame2   S  V  A  I  P  Q  R  S  R  T  R  N  T  I  *  P  F  G  Y  L
Frame1    Q  C  G  H  S  S  E  I  *  N  Q  K  Y  Y  L  T  L  W  V  S
DNA     CAGTGTGGCCATTCCTCAGAGATCTAGAACCAGAAATACTATTTGACCCTTTGGGTATCT 5100
                                             |T/C SNP7, allele1 assoc
                        TCG Boleth           C/T Lucy, C/C Boleth
             SNP3G5063> CCT Lucy
                                                      GCT Boleth
                                                      CCC Lucy <SNP3G5088

Frame3  Y  P  K  E  Y  K  S  F  Y  Y  K  D  T  C  T  R  M  F  T  A
Frame2   P  K  G  I  *  I  I  L  L  *  R  H  M  H  T  Y  V  Y  C  S
Frame1    T  Q  R  N  I  N  H  S  T  I  K  T  H  A  H  V  C  L  L  Q
DNA     ACCCAAAGGAATATAAATCATTCTACTATAAAGACACATGCACACGTATGTTTACTGCAG 5160

Frame3  A  L  F  T  I  A  K  T  W  N  Q  P  K  C  P  S  V  I  D  G
Frame2   T  I  Y  N  S  K  D  L  E  P  T  Q  M  S  I  S  D  R  W  I
Frame1    H  Y  L  Q  *  Q  R  L  G  T  N  P  N  V  H  Q  *  *  M  D
DNA     CACTATTTACAATAGCAAAGACTTGGAACCAACCCAAATGTCCATCAGTGATAGATGGAT 5220

Frame3  *  R  K  C  G  A  Y  H  H  G  I  V  H  S  Q  K  K  E  *  V
Frame2   K  K  M  W  C  I  P  P  W  N  S  T  Q  P  E  K  G  M  S  S
Frame1    K  E  N  V  V  H  T  T  M  E  *  Y  T  A  R  K  R  N  E  F
DNA     AAAGAAAATGTGGTGCATACCACCATGGAATAGTACACAGCCAGAAAAAGGAATGAGTTC 5280

Frame3  H  V  L  C  R  D  M  D  E  A  G  S  H  H  P  Q  Q  T  N  T
Frame2   C  P  L  Q  G  H  G  *  S  W  K  S  S  S  S  A  N  *  H  G
Frame1    M  S  F  A  G  T  W  M  K  L  E  V  I  I  L  S  K  L  T  R
DNA     ATGTCCTTTGCAGGGACATGGATGAAGCTGGAAGTCATCATCCTCAGCAAACTAACACGG 5340

Frame3  G  T  E  N  K  A  P  H  V  L  I  P  K  *  E  L  N  N  D  N
Frame2   N  R  K  Q  S  T  S  C  S  H  S  *  V  R  V  E  Q  *  Q  H
Frame1    E  Q  K  T  K  H  L  M  F  S  F  L  S  E  S  *  T  M  T  T
DNA     GAACAGAAAACAAAGCACCTCATGTTCTCATTCCTAAGT GAGAGTTGAACAATGACAACA 5400

Frame3  T  W  T  Q  G  G  E  Q  H  I  S  G  P  F  G  E  C  G  G  Q
Frame2   M  D  T  G  R  G  T  T  H  I  R  A  F  W  G  V  W  G  A  R
Frame1    H  G  H  R  E  G  N  N  T  Y  Q  G  L  L  G  S  V  G  G  K
DNA     CATGGACACAGGGAGGGGAACAACACATATCAGGGCCTTTTGGGGAGTGTGGGGGGCAAG 5460
              | CT Lucy, C/C Boleth SNPG5407
```

FIGURE 1I

```
Frame3  G T R T * R M G Q * V Q Q T T M A D Y T
Frame2  D E N L E D G S I G A A N H H G R L Y A
Frame1  G R E L R G W V N R C S K P P W Q T I R
DNA     GGACGAGAACTTAGAGGATGGGTCAATAGGTGCAGCAAACCACCATGGC AGACTATACGC 5520
                                                               | TT
        |GG Lucy, GGBoleth
Lucy TT Boleth

Frame3  H V T N L Q V L H M Y P G T * S K I K Q
Frame2  C N K P A G S A H V S W N L K * N K T K
Frame1  M * Q T C R F C T C I L E P K V K * N K
DNA     ATGTAACAAACCTGCAGGTTCTGCACATGTATCCTGGAACCTAAAGTAAAATAAAACAAA 5580

Frame3  S K L K K E S P C L T C M H M F I A A L
Frame2  Q I K K R K P M S Y M Y A Y V H C S T I
Frame1  A N * K K K A H V L H V C I C S L Q H Y
DNA     GCAAATTAAAAAAAGAAAGCCCATGTCTTA CATGTATGCATATGTTCATTGCAGCACTAT 5640

Frame3  F T I A K T W N Q P K C P S M V D W I K
Frame2  H N S K D M E S T * M S I N G R L D K E
Frame1  S Q * Q R H G I N L N V H Q W * T G * R
DNA     TCACAATAGCAAAGACATGGAATCAACCTAAATGTCCATCAATGGTAGACTGGATAAAGA 5700

Frame3  K M W Q I C S T G R I * W R D V S Q K V
Frame2  N V A N M L Y R Q D L M A * C L T E S S
Frame1  K C G K Y A L P A G F D G V M S H R K F
DNA     AAATGTGGCAAATATGCTCTACCGGCAGGATTTGATGGCGTGATGTCTCACAGAAAGTTC 5760

Frame3  L H S Q T W V P R L P A L E A Q Q Q A S
Frame2  P L P D M G P S A S C L G S T A A G I V
Frame1  S T P R H G S L G F L P W K H S S R H R
DNA     TCCACTCCCAGACATGGGTCCCTCGGCTTCCTGCCTTGGAAGCACAGCAGCAGGCATCGT 5820

Frame3  W E G E E L P * G * P I Q A G P P H S L
Frame2  G R * R A S L R M T H P S R S T S Q P S
Frame1  G K V K S F P K D D P S K P V H L T A F
DNA     GGGAAGGTGAAGAGCTTCCCTAAGGATGACCCATCCAAGCCGGTCCACCTCACAGCCTTC 5880

Frame3  P G I Q G W H D P H R A G S R Q A R I Q
Frame2  W D T R L A * P T S C G K S T G Q D P R
Frame1  L G Y K A G M T H I V R E V D R P G S K
DNA     CTGGGATACAAGGCTGGCATGACCCACATCGTGCGGGAAGTCGACAGGCCAGGATCCAAG 5940

Frame3  G E Q E G G G G G C D H C G E A T S G H
Frame2  * T R R R W W R L * P L W R G H Q W A L
Frame1  V N K K E V V E A V T I V E R P P V G I
DNA     GTGAACAAGAAGGAGGTGGTGGAGGCTGTGACCATTGTGGAGAGGCCACCAGTGGGCATT 6000

Frame3  C G L R G N P S R L P D L Q D C L R * A
Frame2  W A A W K P L K A S G L A R L S S L S T
Frame1  V G C V E T P Q G F R T C K T V F A E H
DNA     GTGGGCTGCGTGGAAACCCCTCAAGGCTTCCGGACTTGCAAGACTGTCTTCGCTGAGCAC 6060

Frame3  H Q * * M Q E A F L * E L A * I * E E G
Frame2  S V M N A R G V S I R T G I N L R R R P
Frame1  I S D E C K R R F Y K N W H K S K K K A
DNA     ATCAGTGATGAATGCAAGAGGCGTTTCTATAAGAACTGGCATAAAT CTAAGAAGAAGGCC 6120

Frame3  L Y Q V L Q E M A G * G W Q E A A G E G
Frame2  L P S T A R N G R M R M A R S S W R R T
Frame1  F T K Y C K K W Q D E D G K K Q L E K D
DNA     TTTACCAAGTACTGCAAGAAATGGCAGGATGAGGATGGCAAGAAGCAGCTGGAGAAGGAC 6180
```

FIGURE 1J

```
Frame3  L Q Q H E E V L P S H L R H C P H P D A
Frame2   S A A * R S T A K S S A S L P T P R C T
Frame1   F S S M K K Y C Q V I C V I A H T Q M H  6240
DNA     TTCAGCAGCATGAAGAAGTACTGCCAAGTCATCTGCGTCATTGCCCACACCCAGATGCAC
        ............................................................
Frame3  P A S S V P E E G P P D G D P G E W R H
Frame2   C F L C A R R R P T * W R S R * M E A L
Frame1   L L P L C Q K K A H L M E I Q V N G G T  6300
DNA     CTGCTTC CTCTGTGCCAGAAGAAGGCCCACCTGATGGAGATCCAGGTGAATGGAGGCACT
        ............................................................
Frame3  C G * E A G L G W R E A Q A P G T C E P
Frame2   W L R S W T G L A R G S S T R Y L * T K
Frame1   V A E K L D W A G E R L K H Q V P V N Q  6360
DNA     GTGGCTGAGAAGCTGGACTGGGCTGGCGAGAGGCTCAAGCACCAGGTACCTGTGAACCAA
        ............................................................
Frame3  S V W A G * D D R R H Q G D Q G Q R L Q
Frame2   C L G R M R * S T S S G * P R A K A T K
Frame1   V F G Q D E M I D V I R V T K G K G Y K  6420
DNA     GTGTTTGGGCAGGATGAGATGATCGACGTCATCAGGGTGACCAAGGGCAAAGGCTACAAA
        ............................................................
Frame3  K G H Q S L A H Q E A A P Q D P P R P V
Frame2   G S P V I G T P R S C P A R P T K A C A
Frame1   R V T S H W H T K K L P R K T H Q G L C  6480
DNA     AGGGTCACCAGTCATTGGCACACCAAGAAGCTGCCCCGCAAGACCCACCAAGGCCTGTGC
        ............................................................
Frame3  Q G G L Y W G M A S C S C G L L C G T W
Frame2   R W P V L G H G I L L V W A S L W Y V V
Frame1   K V A C I G A W H P A R V G F S V V R G  6540
DNA     AAGGTGGCCTGTATTGGGGCATGGCATCCTGCTCGTGTGGGCTTCTCTGTGGTACGTGGT
        ............................................................
Frame3  W A E R L P S P H * D Q Q D L * D W L G
Frame2   G R K A T I T A L R S T R S I G L A R A
Frame1   G Q K G Y H H R T E I N K I Y R I G * G  6600
DNA     GGGCAGAAAGGCTACCATCACCGCACTGAGATCAACAAGATCTATAGGA TTGGCTAGGGC
        ............................................................
Frame3  L P Y Q G W Q A D Q E Q C L H * L * P V
Frame2   T L S R M A S * S R T M P P L T M T C L
Frame1   Y L I K D G K L I K N N A S T D Y D L S  6660
DNA     TACCTTATCAAGGATGGCAAGCTGATCAAGAACAATGCCTCCACTGACTATGACCTGTCT
        ............................................................
Frame3  * Q E H Q P F G W L R P L W * S D Q * L
Frame2   T R A S T L W V A S S T M V K * P M T L
Frame1   D K S I N P L G G F V H Y G E V T N D F  6720
DNA     GACAAGAGCATCAACCCTTTGGGTGGCTTCGTCCACTATGGTGAAGTGACCAATGACTTT
        ............................................................
Frame3  C H A E R L C G G N Q E V G A H P P Q V
Frame2   S C * K A V W W E P R S G C S P S T S P
Frame1   V M L K G C V V G T K K W V L T L H K S  6780
DNA     GTCATGCTGAAAGGCTGTGTGGT GGGAACCAAGAAGTGGGTGCTCACCCTCCACAAGTCC
        ............................................................
Frame3  L A G A D K A A G S G E D * P * V H * H
Frame2   C W C R Q S S G L W R R L T L S S L T P
Frame1   L L V Q T K Q R A L E K I D L K F I D T  6840
DNA     TTGCTGGTGCAGACAAAGCAGCGGGCTCTGGAGAAGATTGACCTTAAGTTCATTGACACC
        ............................................................
Frame3  L L Q V W P W P L P D H G G E E S I H G
Frame2   P P S L A M A A S R P R R R R K H S W D
Frame1   S S K F G H G R F Q T T E E K K A F M G  6900
DNA     TCCTCCAAGTTTGGCCATGGCCGCTTCCAGACCACGGAGGAGAAGAAAGCATTCATGGGA
        ............................................................
Frame3  T T Q E R P N C K G R R S L M L G T D I
Frame2   H S R K T E L Q R K K E L N A G N R Y C
Frame1   P L K K D R I A K E E G A * C W E Q I L  6960
DNA     CCACTCAAGAAAGACCGAATTGCAAAGGAAGAAGGAGCTTAATGCTGGGAACAGATATTG
        ............................................................
```

FIGURE 1K

```
Frame3  A  T  G  G  I  S  I  K  V  I  F  H  *  K  K  K  K  K  K  M
Frame2   N  W  W  D  L  N  K  S  Y  F  P  L  K  K  K  E  K  E  N  V
Frame1   Q  L  V  G  S  Q  *  K  L  F  S  I  K  K  K  R  K  R  K  C
DNA     CAACTGGTGGGATCTCAATAAAAGTTATTTTCCATTAAAAAAAAAAGAAAAAGAAAATGT 7020

Frame3  W  H  I  Y  T  T  E  Y  H  A  A  I  K  K  N  E  I  M  S  F
Frame2   A  H  I  H  H  R  I  P  C  S  H  K  K  E  *  D  H  V  L  C
Frame1   G  T  Y  T  P  Q  N  T  M  Q  P  *  K  R  M  R  S  C  P  L
DNA     GGCACATATACACCACAGAATACCATGCAGCCATAAAAAAGAATGAGATCAT GTCCTTTG 7080

Frame3  A  G  T  W  M  E  L  E  A  I  I  L  S  K  L  R  Q  E  Q  K
Frame2   R  N  M  D  G  V  G  G  H  Y  P  *  Q  T  E  A  G  T  E  N
Frame1   Q  E  H  G  W  S  W  R  P  L  S  L  A  N  *  G  R  N  R  K
DNA     CAGGAACATGGATGGAGTTGGAGGCCATTATCCTTAGCAAACTGAGGCAGGAACAGAAAA 7140

Frame3  T  N  Y  H  M  F  S  L  I  S  R  S  Y  M  M  R  T  H  G  H
Frame2   Q  L  P  H  V  L  T  Y  K  *  E  L  Y  D  E  N  T  W  T  H
Frame1   P  I  T  T  C  S  H  L  *  V  G  V  I  *  *  E  H  M  D  T
DNA     CCAATTACCACATGTTCTCACTTATAAGTAGGAGTTATATGATGAGAACACATGGACACA 7200

Frame3  T  E  G  N  N  T  H  W  G  P  L  E  G  R  G  W  E  E  G  E
Frame2   R  R  E  Q  H  T  L  G  S  T  *  G  *  R  V  G  G  G  R
Frame1   Q  K  G  T  T  H  T  G  V  H  L  R  V  E  G  G  R  R  E  R
DNA     CAGAAGGGAACAACACACACTGGGGTCCACTT GAGGGTAGAGGGTGGAGGAGGGAGAGG 7260
```

FIGURE 2

```
SEQ ID NO:3   17  KKKTHPHQKVGKGYKQTLLRGRHLRGQETYEKKLTH ---VYETTDFKQQQYYFHSILGPA 73
                  K   H   K+ K  ++  L    R    Y K +T      + +  DFKQQQ YFHSILGP
SEQ ID NO:4   34  KSYVHSALKIFKTAEECRLDRDEERAYVLYMKYVTVYNLIKKRPDFKQQQDYFHSILGPG 93

SEQ ID NO:3   74  NIKKATGETERLSESLKLRYEEVEIWKKLEEKDXQGEAQWLQQKRQETGREDGSXLAKGS 133
                  NIKKA  E ERLSESLKLRYEE E+ KKLEEKD Q EAQ LQQKRQETGRED G  LAKGS
SEQ ID NO:4   94  NIKKAVEEAERLSESLKLRYEEAEVRKKLEEKDRQEEAQRLQQKRQETGREDGGTLAKGS 153

SEQ ID NO:3  134  LEIVLDSKDKTQKSNGEKNEKCETKEKGAITAKELYTMMMDKNISLIIMXAQRMQXYQDS 193
                  LE VLDSKDKTQKSNGEKNEKCETKEKGAITAKELYTMM DKNISLIIM A+RMQ YQDS
SEQ ID NO:4  154  LENVLDSKDKTQKSNGEKNEKCETKEKGAITAKELYTMMTDKNISLIIMDARRMQDYQDS 213

SEQ ID NO:3  194  CILHSLSVPXXAISPXVTASWIEAHLPYDSIXTWKKRGNVEYMVLLDWFSSAKDLQIXTT 253
                  CILHSLSVP  AISP VTASWIEAHLP DS  TWKKRGNVEY+VLLDWFSSAKDLQI TT
SEQ ID NO:4  214  CILHSLSVPEEAISPGVTASWIEAHLPDDSKDTWKKRGNVEYVVLLDWFSSAKDLQIGTT 273

SEQ ID NO:3  254  LWHLKDXLFKWE ------------KGGYKNWFLCYSQYTTNAKVTPPPQHQNEELSISLD 301
                  L  LKD LFKWE             +GGY+NW LCY QYTTNAKVTPPP+ QNEE+SISLD
SEQ ID NO:4  274  LRSLKDALFKWESKTVLRNEPLVLEGGYENWLLCYPQYTTNAKVTPPPRRQNEEVSISLD 333

SEQ ID NO:3  302  FTYPSLEESIPSKPAAEMPPPPIXVDEDIELISDQISDNDQNERTGPLNISIPVESVAAS 361
                  FTYPSLEESIPSKPAA+ PP   I VDE+IELIS       QNER GPLNIS PVE VAAS
SEQ ID NO:4  334  FTYPSLEESIPSKPAAQTPPAS IEVDENIELIS------GQNERMGPLNISTPVEPVAAS 387

SEQ ID NO:3  362  KSDVSPIIQPVPSIKNVPQIDHTKKLAVKLPEEHIIKSESTNHEQQSPQNEKVIPDCSXK 421
                  KSDVSPIIQPVPSIKNVPQID TKK AVKLPEEH IKSESTNHEQQSPQ+ KVIPD S K
SEQ ID NO:4  388  KSDVSPIIQPVPSIKNVPQIDRTKKPAVKLPEEH RIKSESTNHEQQSPQSGKVIPDRSTK 447

SEQ ID NO:3  422  PVVSSPTLMLTDEEKAHIHAETALLMEKNKQEKELQERQQGKQKE ---TEEGRTRAKSQK 478
                  PVV SPTLMLTDEEKA IHAETALLMEKNKQEKEL+ERQQ +QKE     EE    +AK ++
SEQ ID NO:4  448  PVVFSPTLMLTDEEKARIHAETALLMEKNKQEKELRERQQEEQKEK LRKEEQEQKAKKKQ 507

SEQ ID NO:3  479  E 479
                  E
SEQ ID NO:4  508  E 508
```

FIGURE 3

```
rhodanese             *->tagelkalles.apkliliDvRspefGeryeyegGHIpgAvNvplee
                         ta+el+ ++ +++ +li++ +   +      + y+ ++I ++  vp
SEQ ID NO:3      164     TAKELYTMMMDkNISLIIMXAQRM-----QXYQDSCILHSLSVP-XX 204

rhodanese                eiealldrsgilpdieklhllkdpeelaklfgelgsskdkrvivycrsgr
                         +i+  +++s+i++     hl+ d   ++k+ g+  +       +  + +s +
SEQ ID NO:3      205     AISPXVTASWIEA-----HLPYDSIXTWKKRGNVEY---MVLLDWFSSAK 246

rhodanese                dllrnrrsalaalllkklgypeVyiLkGGykeWlak<-*
                             + +  + ++lk       +  kGGyk+W+ +
SEQ ID NO:3      247     ----DLQIXTTLWHLKDXL---FKWEKGGYKNWFLC     275
```

SEQUENCE LISTING

<110> Applied Research Systems ARS Holding N.V.

<120> NOVEL UBP8rp POLYPEPTIDES AND THEIR USE IN THE TREATMENT OF PSORIASIS

<130> 886 EP

<160> 51

<170> PatentIn version 3.1

<210> 1
<211> 7260
<212> DNA
<213> Homo sapiens

<220>
<221> exon
<222> (851)..(1017)
<223> exon 1

<220>
<221> Intron
<222> (1018)..(1046)
<223>

<220>
<221> exon
<222> (1047)..(1675)
<223> exon 2

<220>
<221> Intron
<222> (1676)..(1718)
<223>

<220>
<221> exon
<222> (1719)..(2371)
<223> exon 3

<220>
<221> allele
<222> (1199)..(1199)
<223> Biallelic marker No. 1

<220>
<221> allele
<222> (1262)..(1262)
<223> Biallelic marker No. 2

<220>
<221> allele
<222> (1426)..(1426)
<223> Biallelic marker No. 3

<220>
<221> allele
<222> (1444)..(1444)
<223> Biallelic marker No. 4

<220>
<221> allele
<222> (1487)..(1487)
<223> Biallelic marker No. 5

<220>
<221> allele
<222> (1490)..(1490)
<223> Biallelic marker No. 6

<220>
<221> allele
<222> (1505)..(1505)
<223> Biallelic marker No. 7

<220>
<221> allele
<222> (1518)..(1518)
<223> Biallelic marker No. 8

<220>
<221> allele
<222> (1554)..(1554)
<223> Biallelic marker No. 9

<220>
<221> allele
<222> (1630)..(1630)

<223> Biallelic marker No. 10

<220>
<221> allele
<222> (1638)..(1638)
<223> Biallelic marker No. 11

<220>
<221> allele
<222> (1680)..(1680)
<223> Biallelic marker No. 12

<220>
<221> allele
<222> (1895)..(1895)
<223> Biallelic marker No. 13

<220>
<221> allele
<222> (2180)..(2180)
<223> Biallelic marker No. 14

<220>
<221> allele
<222> (2449)..(2449)
<223> Biallelic marker No. 15

<220>
<221> allele
<222> (2721)..(2721)
<223> Biallelic marker No. 16

<220>
<221> allele
<222> (3127)..(3127)
<223> Biallelic marker No. 17

<220>
<221> allele
<222> (3137)..(3137)
<223> Biallelic marker No. 18

<220>

<221> allele
<222> (3138)..(3138)
<223> Biallelic marker No. 19

<220>
<221> allele
<222> (3183)..(3183)
<223> Biallelic marker No. 20

<220>
<221> allele
<222> (3222)..(3222)
<223> Biallelic marker No. 21

<220>
<221> allele
<222> (3269)..(3269)
<223> Biallelic marker No. 22

<220>
<221> allele
<222> (3445)..(3445)
<223> Biallelic marker No. 23

<220>
<221> allele
<222> (3470)..(3470)
<223> Biallelic marker No. 24

<220>
<221> allele
<222> (3915)..(3915)
<223> Biallelic marker No. 25

<220>
<221> allele
<222> (3973)..(3973)
<223> Biallelic marker No. 26

<220>
<221> allele
<222> (4254)..(4254)
<223> Biallelic marker No. 27

<220>
<221> allele
<222> (4472)..(4472)
<223> Biallelic marker No. 28

<220>
<221> allele
<222> (4660)..(4660)
<223> Biallelic marker No. 29

<220>
<221> allele
<222> (4770)..(4770)
<223> Biallelic marker No. 30

<220>
<221> allele
<222> (4919)..(4919)
<223> Biallelic marker No. 31

<220>
<221> allele
<222> (4973)..(4973)
<223> Biallelic marker No. 32

<220>
<221> allele
<222> (5063)..(5063)
<223> Biallelic marker No. 33

<220>
<221> allele
<222> (5065)..(5065)
<223> Biallelic marker No. 34

<220>
<221> allele
<222> (5079)..(5079)
<223> Biallelic marker No. 35

<220>
<221> allele
<222> (5080)..(5080)

<223> Biallelic marker No. 36

<220>
<221> allele
<222> (5088)..(5088)
<223> Biallelic marker No. 37

<220>
<221> allele
<222> (5090)..(5090)
<223> Biallelic marker No. 38

<220>
<221> allele
<222> (5407)..(5407)
<223> Biallelic marker No. 39

<220>
<221> allele
<222> (5466)..(5466)
<223> Biallelic marker No. 40

<220>
<221> allele
<222> (5520)..(5520)
<223> Biallelic marker No. 41

<400> 1
tgttatgtag tcttttgttt ggtctctccc ttagcataac gatgtttgag atgatgccat 60

tcattcattt ttgttgctga gcagctgccg agtattgttg gaatcccagt ttattcattg 120

gtttctgtgt ctccagttga tagacatgtg gattcctcca gttagggttt gttattaatg 180

aagccactat aaataactgc ttacaagtgt ggacttacat ttttatttct tttggataaa 240

tacgtatttg tggaattgct gggccatgtg gtaatagatg ggtaactgta taagaaactg 300

ccatacccact ttacaaattg gctgccacat tttttgcatt cctaccagca atatcagaca 360

ttcctatttt ttccatattc ttgccagtgt taagacttat catatgtctt tttaacttta 420

tctgctctag gtgatgtgtg atggtttctc attgtggttt taacttgcac ttctttgatg 480

actagtattg tttgctatct tttcatgttc atctaagcga cttattacat atattttatg 540

```
aactattttg caaattcaat gattaattcc agagactttt tcagaattcc ctagtgtttt     600

ctacatatac aatgaagttg gtgacaaaga aagactttca tttcttcctt tcttatccat     660

tgatcttttt tcttttaaaa ttattattat ttggtagaga tgaggtctca cttatcaggc     720

tggtctcaaa ctcctgatct caagtgatcc tcccacctca gcctcccaaa atgcagggat     780

tacaggcatg agccaccatg cctggtcctt gttgcactgg ttaggatgac tgttaggtgt     840

ttaaacaaga atg atg aga gct cac atg ttt gtt tac aag gaa ctt aaa        889
           Met Met Arg Ala His Met Phe Val Tyr Lys Glu Leu Lys
           1               5                   10

caa att tac aag aaa aaa acc cat ccc cat caa aaa gtg ggc aaa gga       937
Gln Ile Tyr Lys Lys Lys Thr His Pro His Gln Lys Val Gly Lys Gly
    15                  20                  25

tat aaa cag aca ctt ctc aga gga aga cat tta cgt ggc caa gaa aca       985
Tyr Lys Gln Thr Leu Leu Arg Gly Arg His Leu Arg Gly Gln Glu Thr
30                  35                  40                  45

tat gaa aaa aag ctc aca cac gta tat gaa ac  gtgactgttt ataatcctat    1037
Tyr Glu Lys Lys Leu Thr His Val Tyr Glu Thr
                50                  55

ccaaaaaag a act gat ttc aag caa cag cag tat tac ttc cat tca ata      1086
            Thr Asp Phe Lys Gln Gln Gln Tyr Tyr Phe His Ser Ile
                        60                  65

ctt gga cct gca aac atc aaa aaa gcc act gga gaa act gaa cga ctc      1134
Leu Gly Pro Ala Asn Ile Lys Lys Ala Thr Gly Glu Thr Glu Arg Leu
70                  75                  80                  85

tct gaa agc ctt aaa cta aga tat gaa gaa gtt gaa atc tgg aaa aaa      1182
Ser Glu Ser Leu Lys Leu Arg Tyr Glu Glu Val Glu Ile Trp Lys Lys
                90                  95                  100

ctt gag gaa aag gac arg cag ggg gaa gca cag tgg cta caa caa aaa      1230
Leu Glu Glu Lys Asp Xaa Gln Gly Glu Ala Gln Trp Leu Gln Gln Lys
            105                 110                 115

agg cag gaa aca gga aga gag gat ggc agc ayg ttg gct aaa ggt tct      1278
Arg Gln Glu Thr Gly Arg Glu Asp Gly Ser Xaa Leu Ala Lys Gly Ser
        120                 125                 130

ttg gag att gta ttg gat tcc aaa gac aaa acc caa aag agc aat ggt      1326
Leu Glu Ile Val Leu Asp Ser Lys Asp Lys Thr Gln Lys Ser Asn Gly
    135                 140                 145

gaa aag aat gaa aaa tgt gag acc aaa gag aaa gga gca atc aca gca      1374
Glu Lys Asn Glu Lys Cys Glu Thr Lys Glu Lys Gly Ala Ile Thr Ala
```

150 155 160 165

aag gaa cta tac aca atg atg atg gat aaa aac atc agc ttg att ata 1422
Lys Glu Leu Tyr Thr Met Met Met Asp Lys Asn Ile Ser Leu Ile Ile
170 175 180

atg sat gct caa aga atg cag kat tat cag gat tcc tgt att tta cat 1470
Met Xaa Ala Gln Arg Met Gln Xaa Tyr Gln Asp Ser Cys Ile Leu His
185 190 195

tct ctc agt gtt cct gra ara gcc atc agt cca gka gtc act gct agy 1518
Ser Leu Ser Val Pro Xaa Xaa Ala Ile Ser Pro Xaa Val Thr Ala Xaa
200 205 210

tgg att gaa gca cac ctc cca tat gat tct ata gay aca tgg aag aag 1566
Trp Ile Glu Ala His Leu Pro Tyr Asp Ser Ile Asp Thr Trp Lys Lys
215 220 225

agg ggg aat gtg gag tat atg gta ctt ctt gac tgg ttt agt tct gca 1614
Arg Gly Asn Val Glu Tyr Met Val Leu Leu Asp Trp Phe Ser Ser Ala
230 235 240 245

aaa gat tta cag att rga aca acw ctc tgg cat ctg aaa gat gca ctt 1662
Lys Asp Leu Gln Ile Xaa Thr Xaa Leu Trp His Leu Lys Asp Ala Leu
250 255 260

ttc aag tgg gaa a gtaaractgt cctgtgcaat gggccttggg cctttggttt tag 1718
Phe Lys Trp Glu
265

ag gga ggc tat aaa aac tgg ttc ctt tgc tat tcc cag tat aca aca 1765
Lys Gly Gly Tyr Lys Asn Trp Phe Leu Cys Tyr Ser Gln Tyr Thr Thr
270 275 280

aat gct aag gtc act cca ccc cca caa cac cag aat gaa gag ttg tct 1813
Asn Ala Lys Val Thr Pro Pro Pro Gln His Gln Asn Glu Glu Leu Ser
285 290 295

atc tca ttg gat ttt act tat ccc tca ttg gaa gaa tca att cct tct 1861
Ile Ser Leu Asp Phe Thr Tyr Pro Ser Leu Glu Glu Ser Ile Pro Ser
300 305 310

aaa cct gct gcc gag atg cca cct cca cct ata raa gtg gat gaa gac 1909
Lys Pro Ala Ala Glu Met Pro Pro Pro Pro Ile Xaa Val Asp Glu Asp
315 320 325

ata gaa ttg ata agt gat caa ata agt gat aat gat caa aat gag agg 1957
Ile Glu Leu Ile Ser Asp Gln Ile Ser Asp Asn Asp Gln Asn Glu Arg
330 335 340 345

aca gga cca ctg aat ata tca att cca gtt gaa tca gtt gct gct tct 2005
Thr Gly Pro Leu Asn Ile Ser Ile Pro Val Glu Ser Val Ala Ala Ser

350 355 360

aaa tct gat gtt tca ccc atc att cag cca gtg cct agc ata aag aat 2053
Lys Ser Asp Val Ser Pro Ile Ile Gln Pro Val Pro Ser Ile Lys Asn
365 370 375

gtt cca cag att gat cat act aaa aaa ctg gca gtc aaa ttg cct gaa 2101
Val Pro Gln Ile Asp His Thr Lys Lys Leu Ala Val Lys Leu Pro Glu
380 385 390

gag cat ata atc aaa tct gaa agt aca aat cat gag caa cag tct cct 2149
Glu His Ile Ile Lys Ser Glu Ser Thr Asn His Glu Gln Gln Ser Pro
395 400 405

cag aat gaa aaa gtt att cct gat tgt tcc rcc aag cca gta gtt tcc 2197
Gln Asn Glu Lys Val Ile Pro Asp Cys Ser Xaa Lys Pro Val Val Ser
410 415 420 425

tct cca act ctc atg tta aca gat gaa gaa aag gct cat att cat gca 2245
Ser Pro Thr Leu Met Leu Thr Asp Glu Glu Lys Ala His Ile His Ala
430 435 440

gaa act gct ctt cta atg gag aaa aac aaa caa gaa aaa gaa ctt cag 2293
Glu Thr Ala Leu Leu Met Glu Lys Asn Lys Gln Glu Lys Glu Leu Gln
445 450 455

gaa aga cag caa ggg aaa cag aaa gaa act gag gag gga aga aca cga 2341
Glu Arg Gln Gln Gly Lys Gln Lys Glu Thr Glu Glu Gly Arg Thr Arg
460 465 470

gca aaa agc caa aaa gaa aca aga agc tga agaaaatgaa attacacaga 2391
Ala Lys Ser Gln Lys Glu Thr Arg Ser
475 480

agcaacaaaa agcaaaagaa gaaatggaga agaaagaacg tgaacaggcc aagaaagwgg 2451

ataaagaaat ctcagcaaag aagggcaaag aaataacaag agtaaaaaga caaagtaaaa 2511

gtgatcatga aacctctggt gccgagaagt ctgtagagga caggggggaga agatgttcaa 2571

ccccagaagt acagaaaaag tcaacaagag atgtgtccca tacatctgcg acaggggatt 2631

caggttcagg caagcctttt aagattaaag gacaaccaga aactggaatt ctaaggacag 2691

aaacttttag agaggataca gatgatactk aaagaaataa aactcaacga gaaccttcga 2751

taatagcacg aagtgaagaa atggggagga tggtaccagg actgccttca ggctgggcca 2811

agtttcttga tccaatcact ggaacgtttc attattatca ttcaccacta acactgttca 2871

tatgtaccca ctggaaatgg ctccttcatc tgcacctcct tccacccctc caactcataa 2931

aggcaagcca cagattcctg ctaagcagga tagggaacct tccaaactga aatgctctta 2991
ctcctcccca gatataaccc aggctattca agaggaagcc agcagtaact ccaacagtta 3051
atcaggaaga caagccaaca tgctacccta aagctgagat ctcaaggctt tctgcttctc 3111
agatttggaa actcartcct gttttyrgag gttctggacc agctcttact ggacttcgta 3171
acttaggaaa trcttgttat atgaactcaa tattgcagtg cctatgtaat sctccacatt 3231
tggctgatta tttcaaccga aactgttatc aggatgayat taacaagtca aatttgttag 3291
gggcataaag gtgaagtggc agaagaattt ggtataatca tgaaagcccc gtggacagga 3351
cagtatagat atatcagtcc aaaagacctt aaagtcacca ttgggaagat caattaccag 3411
tttgcaggat acagtcaaga ttcacaagaa tttyttctgt tcctaatgga tggtctccrt 3471
gaagatctaa ataaaactga taatcggaag acatataaag aagaaaataa tgatcatctc 3531
aatgacttta aagctgcaga acatgcctgg cagaaacaca agcggctcta tgagtctatt 3591
attgttgcac tttttcaggg tcaattcaaa tctacagtac agtgcctcac ccgtcacaaa 3651
aagtctagga cacttgaggc cttcatgtat ttgtctctac tgatagcatc cacaagtaaa 3711
tgtacattat aggattgcct tagattattt tctaaagaag aaaaactcat agataataac 3771
agattttact gcaatctttg cagagctcga cgggattctt aaaaaagaaa tctggaagtt 3831
accacctgtg cttttagtgc atctgaaaca tttttcctac aatggcaggt ggaaacaaaa 3891
attacagaca tctgtggact tccygttaga aaatcttgcc ttgtcacagt atgttattgg 3951
tccaaagaac aatttgaaga amtataattt gttttctgtt tcagatcact gcggtgggct 4011
ggatggaggc cattacacag cctactgtaa aaatgcagca aaacagcggt ggtttaagtt 4071
tgatgatcat gaagtttctg atatctctgt ttcttctgtg aaatcttcag cagcttatat 4131
cctcttttat acttctttgg gaccatgagt aactgatgta ggcacataag gagacatagg 4191
ttataaacta gttatctttt aaaaggctca gcaacacaat tcttgaaatg cttatcaaga 4251
tartggtagc aatagctggc catttagagg aattctagga cagtgggagc tgtgttacta 4311
gcactatata attcctgtca gtggtgacaa ataacactta acaagtattg cagtaagcat 4371
cacttacagg taccatttat ttcaaaacaa ctttttagt ctgctccaaa gttaaaataa 4431
ttaactagct aagcattatt attctactgg tctaaaaacc wttgtaccct ttttttcctt 4491

ttcactgtta cagcctttc acatttctaa atcccatctt catatactat gaatactcta 4551
gaatgatgtg aagcagatag gaatgtatgt gtacatattt attgcatact tacacatcaa 4611
atcgatatac atagtttaac atgtggtcct ttcgtgaaac ttagaactya gaggattgca 4671
ttttttcttt tgagcatatt ttgagtaact gcagtgcttt cttagggaaa tgacagggca 4731
aagctatttt tctgttggct ttgggggcat ttgggtgcrc taaatcttta tcttaaaaaa 4791
taaatggaaa cttcctttaa ttttttaaaa tgagacatta aaatcttaat gagaaaaatt 4851
taaaaagctc aatatcactg ctcattagag aaatgtaaat caaagccaca atgagatacc 4911
atctccccrcc agtcagaatg gtaattatta aaaagtcaag aaacaataga tgctggtgag 4971
gytgtggaga aataggaaca cttttacact gttgttggga atgtaaacta gttcaaccat 5031
tgtggaagac agtgtggcca ttcctcagag ayckagaacc agaaatayya tttgacscyt 5091
tgggtatcta cccaaaggaa tataaatcat tctactataa agacacatgc acacgtatgt 5151
ttactgcagc actatttaca atagcaaaga cttggaacca acccaaatgt ccatcagtga 5211
tagatggata aagaaaatgt ggtgcatacc accatggaat agtacacagc cagaaaaagg 5271
aatgagttca tgtcctttgc agggacatgg atgaagctgg aagtcatcat cctcagcaaa 5331
ctaacacggg aacagaaaac aaagcacctc atgttctcat tcctaagtga gagttgaaca 5391
atgacaacac atggayacag ggaggggaac aacacatatc agggcctttt ggggagtgtg 5451
gggggcaagg gacgrgaact tagaggatgg gtcaataggt gcagcaaacc accatggcag 5511
actatacgya tgtaacaaac ctgcaggttc tgcacatgta tcctggaacc taaagtaaaa 5571
taaaacaaag caaattaaaa aaagaaagcc catgtcttac atgtatgcat atgttcattg 5631
cagcactatt cacaatagca aagacatgga atcaacctaa atgtccatca atggtagact 5691
ggataaagaa aatgtggcaa atatgctcta ccggcaggat ttgatggcgt gatgtctcac 5751
agaaagttct ccactcccag acatgggtcc ctcggcttcc tgccttggaa gcacagcagc 5811
aggcatcgtg ggaaggtgaa gagcttccct aaggatgacc catccaagcc ggtccacctc 5871
acagccttcc tgggatacaa ggctggcatg acccacatcg tgcgggaagt cgacaggcca 5931
ggatccaagg tgaacaagaa ggaggtggtg gaggctgtga ccattgtgga gaggccacca 5991
gtgggcattg tgggctgcgt ggaaacccct caaggcttcc ggacttgcaa gactgtcttc 6051

gctgagcaca tcagtgatga atgcaagagg cgtttctata agaactggca taaatctaag 6111
aagaaggcct ttaccaagta ctgcaagaaa tggcaggatg aggatggcaa gaagcagctg 6171
gagaaggact tcagcagcat gaagaagtac tgccaagtca tctgcgtcat tgcccacacc 6231
cagatgcacc tgcttcctct gtgccagaag aaggcccacc tgatggagat ccaggtgaat 6291
ggaggcactg tggctgagaa gctggactgg gctggcgaga ggctcaagca ccaggtacct 6351
gtgaaccaag tgtttgggca ggatgagatg atcgacgtca tcagggtgac caagggcaaa 6411
ggctacaaaa gggtcaccag tcattggcac accaagaagc tgccccgcaa gacccaccaa 6471
ggcctgtgca aggtggcctg tattggggca tggcatcctg ctcgtgtggg cttctctgtg 6531
gtacgtggtg ggcagaaagg ctaccatcac cgcactgaga tcaacaagat ctataggatt 6591
ggctagggct accttatcaa ggatggcaag ctgatcaaga acaatgcctc cactgactat 6651
gacctgtctg acaagagcat caaccctttg ggtggcttcg tccactatgg tgaagtgacc 6711
aatgactttg tcatgctgaa aggctgtgtg gtgggaacca agaagtgggt gctcaccctc 6771
cacaagtcct tgctggtgca gacaaagcag cgggctctgg agaagattga ccttaagttc 6831
attgacacct cctccaagtt tggccatggc cgcttccaga ccacggagga gaagaaagca 6891
ttcatgggac cactcaagaa agaccgaatt gcaaaggaag aaggagctta atgctgggaa 6951
cagatattgc aactggtggg atctcaataa aagttatttt ccattaaaaa aaaaagaaaa 7011
agaaaatgtg gcacatatac accacagaat accatgcagc cataaaaaag aatgagatca 7071
tgtcctttgc aggaacatgg atggagttgg aggccattat ccttagcaaa ctgaggcagg 7131
aacagaaaac caattaccac atgttctcac ttataagtag gagttatatg atgagaacac 7191
atggacacac agaagggaac aacacacact ggggtccact tgagggtaga gggtgggagg 7251
agggagagg 7260

<210> 2
<211> 1449
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(1449)
<223>

<220>
<221> misc_feature
<222> (1)..(167)
<223> exon 1

<220>
<221> misc_feature
<222> (168)..(796)
<223> exon 2

<220>
<221> misc_feature
<222> (797)..(1449)
<223> exon 3

<400> 2
```
atg atg aga gct cac atg ttt gtt tac aag gaa ctt aaa caa att tac      48
Met Met Arg Ala His Met Phe Val Tyr Lys Glu Leu Lys Gln Ile Tyr
1               5                   10                  15

aag aaa aaa acc cat ccc cat caa aaa gtg ggc aaa gga tat aaa cag      96
Lys Lys Lys Thr His Pro His Gln Lys Val Gly Lys Gly Tyr Lys Gln
            20                  25                  30

aca ctt ctc aga gga aga cat tta cgt ggc caa gaa aca tat gaa aaa     144
Thr Leu Leu Arg Gly Arg His Leu Arg Gly Gln Glu Thr Tyr Glu Lys
        35                  40                  45

aag ctc aca cac gta tat gaa aca act gat ttc aag caa cag cag tat     192
Lys Leu Thr His Val Tyr Glu Thr Thr Asp Phe Lys Gln Gln Gln Tyr
    50                  55                  60

tac ttc cat tca ata ctt gga cct gca aac atc aaa aaa gcc act gga     240
Tyr Phe His Ser Ile Leu Gly Pro Ala Asn Ile Lys Lys Ala Thr Gly
65                  70                  75                  80

gaa act gaa cga ctc tct gaa agc ctt aaa cta aga tat gaa gaa gtt     288
Glu Thr Glu Arg Leu Ser Glu Ser Leu Lys Leu Arg Tyr Glu Glu Val
                85                  90                  95

gaa atc tgg aaa aaa ctt gag gaa aag gac arg cag ggg gaa gca cag     336
Glu Ile Trp Lys Lys Leu Glu Glu Lys Asp Xaa Gln Gly Glu Ala Gln
            100                 105                 110

tgg cta caa caa aaa agg cag gaa aca gga aga gag gat ggc agc ayg     384
Trp Leu Gln Gln Lys Arg Gln Glu Thr Gly Arg Glu Asp Gly Ser Xaa
        115                 120                 125
```

ttg gct aaa ggt tct ttg gag att gta ttg gat tcc aaa gac aaa acc 432
Leu Ala Lys Gly Ser Leu Glu Ile Val Leu Asp Ser Lys Asp Lys Thr
130 135 140

caa aag agc aat ggt gaa aag aat gaa aaa tgt gag acc aaa gag aaa 480
Gln Lys Ser Asn Gly Glu Lys Asn Glu Lys Cys Glu Thr Lys Glu Lys
145 150 155 160

gga gca atc aca gca aag gaa cta tac aca atg atg atg gat aaa aac 528
Gly Ala Ile Thr Ala Lys Glu Leu Tyr Thr Met Met Met Asp Lys Asn
165 170 175

atc agc ttg att ata atg sat gct caa aga atg cag kat tat cag gat 576
Ile Ser Leu Ile Ile Met Xaa Ala Gln Arg Met Gln Xaa Tyr Gln Asp
180 185 190

tcc tgt att tta cat tct ctc agt gtt cct gra ara gcc atc agt cca 624
Ser Cys Ile Leu His Ser Leu Ser Val Pro Xaa Xaa Ala Ile Ser Pro
195 200 205

gka gtc act gct agy tgg att gaa gca cac ctc cca tat gat tct ata 672
Xaa Val Thr Ala Xaa Trp Ile Glu Ala His Leu Pro Tyr Asp Ser Ile
210 215 220

gay aca tgg aag aag agg ggg aat gtg gag tat atg gta ctt ctt gac 720
Asp Thr Trp Lys Lys Arg Gly Asn Val Glu Tyr Met Val Leu Leu Asp
225 230 235 240

tgg ttt agt tct gca aaa gat tta cag att rga aca acw ctc tgg cat 768
Trp Phe Ser Ser Ala Lys Asp Leu Gln Ile Xaa Thr Xaa Leu Trp His
245 250 255

ctg aaa gat gca ctt ttc aag tgg gaa aag gga ggc tat aaa aac tgg 816
Leu Lys Asp Ala Leu Phe Lys Trp Glu Lys Gly Gly Tyr Lys Asn Trp
260 265 270

ttc ctt tgc tat tcc cag tat aca aca aat gct aag gtc act cca ccc 864
Phe Leu Cys Tyr Ser Gln Tyr Thr Thr Asn Ala Lys Val Thr Pro Pro
275 280 285

cca caa cac cag aat gaa gag ttg tct atc tca ttg gat ttt act tat 912
Pro Gln His Gln Asn Glu Glu Leu Ser Ile Ser Leu Asp Phe Thr Tyr
290 295 300

ccc tca ttg gaa gaa tca att cct tct aaa cct gct gcc gag atg cca 960
Pro Ser Leu Glu Glu Ser Ile Pro Ser Lys Pro Ala Ala Glu Met Pro
305 310 315 320

cct cca cct ata raa gtg gat gaa gac ata gaa ttg ata agt gat caa 1008
Pro Pro Pro Ile Xaa Val Asp Glu Asp Ile Glu Leu Ile Ser Asp Gln
325 330 335

```
ata agt gat aat gat caa aat gag agg aca gga cca ctg aat ata tca      1056
Ile Ser Asp Asn Asp Gln Asn Glu Arg Thr Gly Pro Leu Asn Ile Ser
                340                     345                     350

att cca gtt gaa tca gtt gct gct tct aaa tct gat gtt tca ccc atc      1104
Ile Pro Val Glu Ser Val Ala Ala Ser Lys Ser Asp Val Ser Pro Ile
            355                     360                     365

att cag cca gtg cct agc ata aag aat gtt cca cag att gat cat act      1152
Ile Gln Pro Val Pro Ser Ile Lys Asn Val Pro Gln Ile Asp His Thr
        370                     375                     380

aaa aaa ctg gca gtc aaa ttg cct gaa gag cat ata atc aaa tct gaa      1200
Lys Lys Leu Ala Val Lys Leu Pro Glu Glu His Ile Ile Lys Ser Glu
385                     390                     395                     400

agt aca aat cat gag caa cag tct cct cag aat gaa aaa gtt att cct      1248
Ser Thr Asn His Glu Gln Gln Ser Pro Gln Asn Glu Lys Val Ile Pro
                405                     410                     415

gat tgt tcc rcc aag cca gta gtt tcc tct cca act ctc atg tta aca      1296
Asp Cys Ser Xaa Lys Pro Val Val Ser Ser Pro Thr Leu Met Leu Thr
                420                     425                     430

gat gaa gaa aag gct cat att cat gca gaa act gct ctt cta atg gag      1344
Asp Glu Glu Lys Ala His Ile His Ala Glu Thr Ala Leu Leu Met Glu
            435                     440                     445

aaa aac aaa caa gaa aaa gaa ctt cag gaa aga cag caa ggg aaa cag      1392
Lys Asn Lys Gln Glu Lys Glu Leu Gln Glu Arg Gln Gln Gly Lys Gln
        450                     455                     460

aaa gaa act gag gag gga aga aca cga gca aaa agc caa aaa gaa aca      1440
Lys Glu Thr Glu Glu Gly Arg Thr Arg Ala Lys Ser Gln Lys Glu Thr
465                     470                     475                     480

aga agc tga                                                          1449
Arg Ser


&lt;210&gt;  3
&lt;211&gt;  482
&lt;212&gt;  PRT
&lt;213&gt;  Homo sapiens

&lt;220&gt;
&lt;221&gt;  misc_feature
&lt;222&gt;  (107)..(107)
&lt;223&gt;  The 'Xaa' at location 107 stands for Arg, or Lys.

&lt;220&gt;
```

<221> misc_feature
<222> (128)..(128)
<223> The 'Xaa' at location 128 stands for Thr, or Met.

<220>
<221> misc_feature
<222> (183)..(183)
<223> The 'Xaa' at location 183 stands for Asp, or His.

<220>
<221> misc_feature
<222> (189)..(189)
<223> The 'Xaa' at location 189 stands for Asp, or Tyr.

<220>
<221> misc_feature
<222> (203)..(203)
<223> The 'Xaa' at location 203 stands for Gly, or Glu.

<220>
<221> misc_feature
<222> (204)..(204)
<223> The 'Xaa' at location 204 stands for Arg, or Lys.

<220>
<221> misc_feature
<222> (209)..(209)
<223> The 'Xaa' at location 209 stands for Gly, or Val.

<220>
<221> misc_feature
<222> (213)..(213)
<223> The 'Xaa' at location 213 stands for Ser.

<220>
<221> misc_feature
<222> (251)..(251)
<223> The 'Xaa' at location 251 stands for Gly, or Arg.

<220>
<221> misc_feature
<222> (253)..(253)
<223> The 'Xaa' at location 253 stands for Thr.

<220>
<221> misc_feature
<222> (325)..(325)
<223> The 'Xaa' at location 325 stands for Glu, or Lys.

<220>
<221> misc_feature
<222> (420)..(420)

<223> The 'Xaa' at location 420 stands for Ala, or Thr.

<400> 3

```
Met Met Arg Ala His Met Phe Val Tyr Lys Glu Leu Lys Gln Ile Tyr
1               5                   10                  15

Lys Lys Lys Thr His Pro His Gln Lys Val Gly Lys Gly Tyr Lys Gln
            20                  25                  30

Thr Leu Leu Arg Gly Arg His Leu Arg Gly Gln Glu Thr Tyr Glu Lys
        35                  40                  45

Lys Leu Thr His Val Tyr Glu Thr Thr Asp Phe Lys Gln Gln Gln Tyr
    50                  55                  60

Tyr Phe His Ser Ile Leu Gly Pro Ala Asn Ile Lys Lys Ala Thr Gly
65                  70                  75                  80

Glu Thr Glu Arg Leu Ser Glu Ser Leu Lys Leu Arg Tyr Glu Glu Val
                85                  90                  95

Glu Ile Trp Lys Lys Leu Glu Glu Lys Asp Xaa Gln Gly Glu Ala Gln
            100                 105                 110

Trp Leu Gln Gln Lys Arg Gln Glu Thr Gly Arg Glu Asp Gly Ser Xaa
        115                 120                 125

Leu Ala Lys Gly Ser Leu Glu Ile Val Leu Asp Ser Lys Asp Lys Thr
    130                 135                 140

Gln Lys Ser Asn Gly Glu Lys Asn Glu Lys Cys Glu Thr Lys Glu Lys
145                 150                 155                 160

Gly Ala Ile Thr Ala Lys Glu Leu Tyr Thr Met Met Met Asp Lys Asn
                165                 170                 175

Ile Ser Leu Ile Ile Met Xaa Ala Gln Arg Met Gln Xaa Tyr Gln Asp
            180                 185                 190
```

Ser Cys Ile Leu His Ser Leu Ser Val Pro Xaa Xaa Ala Ile Ser Pro
195 200 205

Xaa Val Thr Ala Xaa Trp Ile Glu Ala His Leu Pro Tyr Asp Ser Ile
210 215 220

Asp Thr Trp Lys Lys Arg Gly Asn Val Glu Tyr Met Val Leu Leu Asp
225 230 235 240

Trp Phe Ser Ser Ala Lys Asp Leu Gln Ile Xaa Thr Xaa Leu Trp His
245 250 255

Leu Lys Asp Ala Leu Phe Lys Trp Glu Lys Gly Gly Tyr Lys Asn Trp
260 265 270

Phe Leu Cys Tyr Ser Gln Tyr Thr Thr Asn Ala Lys Val Thr Pro Pro
275 280 285

Pro Gln His Gln Asn Glu Glu Leu Ser Ile Ser Leu Asp Phe Thr Tyr
290 295 300

Pro Ser Leu Glu Glu Ser Ile Pro Ser Lys Pro Ala Ala Glu Met Pro
305 310 315 320

Pro Pro Pro Ile Xaa Val Asp Glu Asp Ile Glu Leu Ile Ser Asp Gln
325 330 335

Ile Ser Asp Asn Asp Gln Asn Glu Arg Thr Gly Pro Leu Asn Ile Ser
340 345 350

Ile Pro Val Glu Ser Val Ala Ala Ser Lys Ser Asp Val Ser Pro Ile
355 360 365

Ile Gln Pro Val Pro Ser Ile Lys Asn Val Pro Gln Ile Asp His Thr
370 375 380

Lys Lys Leu Ala Val Lys Leu Pro Glu Glu His Ile Ile Lys Ser Glu
385 390 395 400

Ser Thr Asn His Glu Gln Gln Ser Pro Gln Asn Glu Lys Val Ile Pro
405 410 415

Asp Cys Ser Xaa Lys Pro Val Val Ser Ser Pro Thr Leu Met Leu Thr
420 425 430

Asp Glu Glu Lys Ala His Ile His Ala Glu Thr Ala Leu Leu Met Glu
435 440 445

Lys Asn Lys Gln Glu Lys Glu Leu Gln Glu Arg Gln Gln Gly Lys Gln
450 455 460

Lys Glu Thr Glu Glu Gly Arg Thr Arg Ala Lys Ser Gln Lys Glu Thr
465 470 475 480

Arg Ser

<210> 4
<211> 1118
<212> PRT
<213> Homo sapiens

<400> 4

Met Pro Ala Val Ala Ser Val Pro Lys Glu Leu Tyr Leu Ser Ser Ser
1 5 10 15

Leu Lys Asp Leu Asn Lys Lys Thr Glu Val Lys Pro Glu Lys Ile Ser
20 25 30

Thr Lys Ser Tyr Val His Ser Ala Leu Lys Ile Phe Lys Thr Ala Glu
35 40 45

Glu Cys Arg Leu Asp Arg Asp Glu Glu Arg Ala Tyr Val Leu Tyr Met
50 55 60

Lys Tyr Val Thr Val Tyr Asn Leu Ile Lys Lys Arg Pro Asp Phe Lys
65 70 75 80

Gln Gln Gln Asp Tyr Phe His Ser Ile Leu Gly Pro Gly Asn Ile Lys

85 90 95

Lys Ala Val Glu Glu Ala Glu Arg Leu Ser Glu Ser Leu Lys Leu Arg
100 105 110

Tyr Glu Glu Ala Glu Val Arg Lys Lys Leu Glu Glu Lys Asp Arg Gln
115 120 125

Glu Glu Ala Gln Arg Leu Gln Gln Lys Arg Gln Glu Thr Gly Arg Glu
130 135 140

Asp Gly Gly Thr Leu Ala Lys Gly Ser Leu Glu Asn Val Leu Asp Ser
145 150 155 160

Lys Asp Lys Thr Gln Lys Ser Asn Gly Glu Lys Asn Glu Lys Cys Glu
165 170 175

Thr Lys Glu Lys Gly Ala Ile Thr Ala Lys Glu Leu Tyr Thr Met Met
180 185 190

Thr Asp Lys Asn Ile Ser Leu Ile Ile Met Asp Ala Arg Arg Met Gln
195 200 205

Asp Tyr Gln Asp Ser Cys Ile Leu His Ser Leu Ser Val Pro Glu Glu
210 215 220

Ala Ile Ser Pro Gly Val Thr Ala Ser Trp Ile Glu Ala His Leu Pro
225 230 235 240

Asp Asp Ser Lys Asp Thr Trp Lys Lys Arg Gly Asn Val Glu Tyr Val
245 250 255

Val Leu Leu Asp Trp Phe Ser Ser Ala Lys Asp Leu Gln Ile Gly Thr
260 265 270

Thr Leu Arg Ser Leu Lys Asp Ala Leu Phe Lys Trp Glu Ser Lys Thr
275 280 285

Val Leu Arg Asn Glu Pro Leu Val Leu Glu Gly Gly Tyr Glu Asn Trp

290 295 300

Leu Leu Cys Tyr Pro Gln Tyr Thr Thr Asn Ala Lys Val Thr Pro Pro
305 310 315 320

Pro Arg Arg Gln Asn Glu Glu Val Ser Ile Ser Leu Asp Phe Thr Tyr
325 330 335

Pro Ser Leu Glu Glu Ser Ile Pro Ser Lys Pro Ala Ala Gln Thr Pro
340 345 350

Pro Ala Ser Ile Glu Val Asp Glu Asn Ile Glu Leu Ile Ser Gly Gln
355 360 365

Asn Glu Arg Met Gly Pro Leu Asn Ile Ser Thr Pro Val Glu Pro Val
370 375 380

Ala Ala Ser Lys Ser Asp Val Ser Pro Ile Ile Gln Pro Val Pro Ser
385 390 395 400

Ile Lys Asn Val Pro Gln Ile Asp Arg Thr Lys Lys Pro Ala Val Lys
405 410 415

Leu Pro Glu Glu His Arg Ile Lys Ser Glu Ser Thr Asn His Glu Gln
420 425 430

Gln Ser Pro Gln Ser Gly Lys Val Ile Pro Asp Arg Ser Thr Lys Pro
435 440 445

Val Val Phe Ser Pro Thr Leu Met Leu Thr Asp Glu Glu Lys Ala Arg
450 455 460

Ile His Ala Glu Thr Ala Leu Leu Met Glu Lys Asn Lys Gln Glu Lys
465 470 475 480

Glu Leu Arg Glu Arg Gln Gln Glu Glu Gln Lys Glu Lys Leu Arg Lys
485 490 495

Glu Glu Gln Glu Gln Lys Ala Lys Lys Lys Gln Glu Ala Glu Glu Asn

500 505 510

Glu Ile Thr Glu Lys Gln Gln Lys Ala Lys Glu Glu Met Glu Lys Lys
515 520 525

Glu Ser Glu Gln Ala Lys Lys Glu Asp Lys Glu Thr Ser Ala Lys Arg
530 535 540

Gly Lys Glu Ile Thr Gly Val Lys Arg Gln Ser Lys Ser Glu His Glu
545 550 555 560

Thr Ser Asp Ala Lys Lys Ser Val Glu Asp Arg Gly Lys Arg Cys Pro
565 570 575

Thr Pro Glu Ile Gln Lys Lys Ser Thr Gly Asp Val Pro His Thr Ser
580 585 590

Val Thr Gly Asp Ser Gly Ser Gly Lys Pro Phe Lys Ile Lys Gly Gln
595 600 605

Pro Glu Ser Gly Ile Leu Arg Thr Gly Thr Phe Arg Glu Asp Thr Asp
610 615 620

Asp Thr Glu Arg Asn Lys Ala Gln Arg Glu Pro Leu Thr Arg Ala Arg
625 630 635 640

Ser Glu Glu Met Gly Arg Ile Val Pro Gly Leu Pro Ser Gly Trp Ala
645 650 655

Lys Phe Leu Asp Pro Ile Thr Gly Thr Phe Arg Tyr Tyr His Ser Pro
660 665 670

Thr Asn Thr Val His Met Tyr Pro Pro Glu Met Ala Pro Ser Ser Ala
675 680 685

Pro Pro Ser Thr Pro Pro Thr His Lys Ala Lys Pro Gln Ile Pro Ala
690 695 700

Glu Arg Asp Arg Glu Pro Ser Lys Leu Lys Arg Ser Tyr Ser Ser Pro

705 710 715 720

Asp Ile Thr Gln Ala Ile Gln Glu Glu Glu Lys Arg Lys Pro Thr Val
725 730 735

Thr Pro Thr Val Asn Arg Glu Asn Lys Pro Thr Cys Tyr Pro Lys Ala
740 745 750

Glu Ile Ser Arg Leu Ser Ala Ser Gln Ile Arg Asn Leu Asn Pro Val
755 760 765

Phe Gly Gly Ser Gly Pro Ala Leu Thr Gly Leu Arg Asn Leu Gly Asn
770 775 780

Thr Cys Tyr Met Asn Ser Ile Leu Gln Cys Leu Cys Asn Ala Pro His
785 790 795 800

Leu Ala Asp Tyr Phe Asn Arg Asn Cys Tyr Gln Asp Asp Ile Asn Arg
805 810 815

Ser Asn Leu Leu Gly His Lys Gly Glu Val Ala Glu Glu Phe Gly Ile
820 825 830

Ile Met Lys Ala Leu Trp Thr Gly Gln Tyr Arg Tyr Ile Ser Pro Lys
835 840 845

Asp Phe Lys Ile Thr Ile Gly Lys Ile Asn Asp Gln Phe Ala Gly Tyr
850 855 860

Ser Gln Gln Asp Ser Gln Glu Leu Leu Leu Phe Leu Met Asp Gly Leu
865 870 875 880

His Glu Asp Leu Asn Lys Ala Asp Asn Arg Lys Arg Tyr Lys Glu Glu
885 890 895

Asn Asn Asp His Leu Asp Asp Phe Lys Ala Ala Glu His Ala Trp Gln
900 905 910

Lys His Lys Gln Leu Asn Glu Ser Ile Ile Val Ala Leu Phe Gln Gly

915 920 925

Gln Phe Lys Ser Thr Val Gln Cys Leu Thr Cys His Lys Lys Ser Arg
930 935 940

Thr Phe Glu Ala Phe Met Tyr Leu Ser Leu Pro Leu Ala Ser Thr Ser
945 950 955 960

Lys Cys Thr Leu Gln Asp Cys Leu Arg Leu Phe Ser Lys Glu Glu Lys
965 970 975

Leu Thr Asp Asn Asn Arg Phe Tyr Cys Ser His Cys Arg Ala Arg Arg
980 985 990

Asp Ser Leu Lys Lys Ile Glu Ile Trp Lys Leu Pro Pro Val Leu Leu
995 1000 1005

Val His Leu Lys Arg Phe Ser Tyr Asp Gly Arg Trp Lys Gln Lys
1010 1015 1020

Leu Gln Thr Ser Val Asp Phe Pro Leu Glu Asn Leu Asp Leu Ser
1025 1030 1035

Gln Tyr Val Ile Gly Pro Lys Asn Asn Leu Lys Lys Tyr Asn Leu
1040 1045 1050

Phe Ser Val Ser Asn His Tyr Gly Gly Leu Asp Gly Gly His Tyr
1055 1060 1065

Thr Ala Tyr Cys Lys Asn Ala Ala Arg Gln Arg Trp Phe Lys Phe
1070 1075 1080

Asp Asp His Glu Val Ser Asp Ile Ser Val Ser Ser Val Lys Ser
1085 1090 1095

Ser Ala Ala Tyr Ile Leu Phe Tyr Thr Ser Leu Gly Pro Arg Val
1100 1105 1110

Thr Asp Val Ala Thr

1115

<210> 5
<211> 29
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 5
atgatgagag ctcacatgtt tgtttacaa 29

<210> 6
<211> 24
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 6
tttcaagaat tgtgttgctg agcc 24

<210> 7
<211> 30
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 7
gctcacatgt ttgtttacaa ggaacttaaa 30

<210> 8
<211> 26
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 8

tgcctacatc agttactcat ggtccc 26

<210> 9
<211> 22
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 9
gtttgtttac aaggaactta aa 22

<210> 10
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 10
gatttcaagc aacagcagt 19

<210> 11
<211> 20
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 11
gaaagatgca cttttcaagt 20

<210> 12
<211> 22
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 12

aagagttgtc tatctcattg ga 22

<210> 13
<211> 22
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 13
gaccatgagt aactgatgta gg 22

<210> 14
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 14
catggtccca aagaagtat 19

<210> 15
<211> 17
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 15
aaaccaccgc tgttttg 17

<210> 16
<211> 21
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 16

tgaccttagc atttgttgta t 21

<210> 17
<211> 20
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 17
cttgaaaagt gcatctttca 20

<210> 18
<211> 18
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 18
ctgcctgtcc ttttcctc 18

<210> 19
<211> 21
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 19
gtgtgagctt tttttcatat g 21

<210> 20
<211> 20
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 20

cgtaaatgtc ttcctctgag 20

<210> 21
<211> 30
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 21
aagtgataat gatcaaaatg agaggacagg 30

<210> 22
<211> 24
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 22
tgttccacca agccagtagt ttcc 24

<210> 23
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 23
acatcaaaaa agccactgg 19

<210> 24
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 24

cactgctagc tggattgaa 19

<210> 25
<211> 22
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 25
cctcccagat gattctatag at 22

<210> 26
<211> 22
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 26
cacctataaa agtggatgaa ga 22

<210> 27
<211> 21
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 27
cagtagtttc ctctccaact c 21

<210> 28
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 28

gaggagggaa gaacacgag 19

<210> 29
<211> 26
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 29
gatacttaaa gaaataaaac tcaacg 26

<210> 30
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 30
ttccaaactg aaatgctct 19

<210> 31
<211> 17
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 31
tgcccttctt tgctgag 17

<210> 32
<211> 22
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 32

tgactgccag ttttttagta tg 22

<210> 33
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 33
atctttcaga tgccagagt 19

<210> 34
<211> 18
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 34
cctttagcca acatgctg 18

<210> 35
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 35
tcgttcagtt tctccagtg 19

<210> 36
<211> 22
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 36

cgagatgcca cctccaccta ta 22

<210> 37
<211> 23
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 37
tcagtggtcc tgtcctctca ttt 23

<210> 38
<211> 23
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 38
tgatcaaata agtgataatg atc 23

<210> 39
<211> 26
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 39
aattgatata ttcagtggtc ctgtcc 26

<210> 40
<211> 21
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 40

cgagatgcca cctccaccta t 21

<210> 41
<211> 21
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 41
tctcattttg atcattatca c 21

<210> 42
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 42
agacaagcca acatgctac 19

<210> 43
<211> 21
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 43
tggtgacttt aaggtctttt g 21

<210> 44
<211> 20
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 44

ctcacccgtc acaaaaagtc 20

<210> 45
<211> 20
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 45
gcttcttgcc atcctcatcc 20

<210> 46
<211> 19
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 46
cctgtgcttt tagtgcatc 19

<210> 47
<211> 20
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 47
gctgaagatt tcacagaaga 20

<210> 48
<211> 22
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 48

ctttcgtgaa acttagaact ca 22

<210> 49
<211> 20
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 49
cattgtggct ttgatttaca 20

<210> 50
<211> 20
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 50
gaaacaatag atgctggtga 20

<210> 51
<211> 15
<212> DNA
<213> Artificial

<220>
<221> source
<223> /note="Description of artificial sequence: primer"

<400> 51
actgatggac atttg 15